

INDEX

	PAGE
AMERIO, LUIGI. Sul calcolo delle soluzioni dei problemi al contorno per la equazioni lineari del secondo ordine di tipo ellittico,	447
ARTIN, EMIL. The free product of groups,	1
BAER, REINHOLD. The double chain condition in cyclic operator groups,	37
BAER, REINHOLD. Projectivities of finite projective planes,	653
BARLAZ, JOSHUA. On some triangular summability methods,	139
BATES, GRACE E. Free loops and nets and their generalizations,	499
BERNHART, ARTHUR. Six-rings in minimal five-color maps,	391
CHING, R. H. Skew sets,	493
CHAUVER, RICHARD. On the zeta-functions of algebraic number fields,	243
CHAUVER, RICHARD. Applications of induced characters,	709
CHEN, BAILEY and NEAL H. MCCOY. Radicals and subdirect sums,	46
CHEN, ELLEN F. and R. C. BUCK. A note on finitely-additive measures	413
BUCK, R. C. See Ellen F. Buck and R. C. Buck, page 413.	
BUSEMANN, HERBERT. The isoperimetric problem in the Minkowski plane,	863
CHUNG, KAI-LAI. Note on some lattice numbers,	189
CROSBY, W. J. R. Generic algebras,	333
DE CICCIO, JOHN. Union-preserving transformations of higher order surface-elements,	104
DE CICCIO, JOHN. See Kasner and De Ciccio, page 575.	
DOWKER, C. H. Mapping theorems for non-compact spaces,	200
ESSER, MARTINUS. Analyticity in Hilbert space and self-adjoint transformations,	825
HABBOUR, M. N. On convergence of the product of basic sets of polynomials,	583
GODEAUX, LUCIEN. Sur une suite de quadriques associée à un congruence W ,	490
GROVE, V. G. On congruences and conjugate nets,	59
HARTMAN, PHILIP. On the ergodic theorems,	193
HARTMAN, PHILIP. Systems of total differential equations and Liouville's theorem on conformal mappings,	327

	PAGE
HARTMAN, PHILIP. Tauber's theorem and absolute constants, . . .	599
HARTMAN, PHILIP and AUREL WINTNER. On Möbius' inversion, . . .	853
HAVILAND, E. K. A note on Laplace transforms of functions whose spectra are confined to a given set,	279
HSIUNG, CHUAN-CHIH. A study on the theory of conjugate nets, . . .	379
HSIUNG, CHUAN-CHIH. Projective theory of surfaces and conjugate nets in four-dimensional space,	607
JACOBSON, N. A note on division rings,	27
KAGNO, I. N. Desargues' and Pappus' graphs and their groups, . . .	859
KAGNO, I. N. Corrections,	872
KAPLANSKY, IRVING. Topological rings,	153
KASNER, EDWARD and JOHN DE CICCIO. Extensions of harmonic transformations,	575
KELLEY, J. L. Simple links and fixed sets under continuous mappings, . . .	348
LEE, HWA-CHUNG. On skew-metric spaces and function groups, . . .	790
LEWIS, D. C. Polynomial least square approximations,	273
MANGILL, JULIAN D. Unilateral variations with variable end-points, . . .	121
McCOY, NEAL H. See Brown and McCoy, page 46.	
MESERVE, BRUCE ELWYN. Inequalities of higher degree in one unknown,	357
NEHARI, ZEEV. The elliptic modular function and a class of analytic functions first considered by Hurwitz,	70
OWENS, O. G. Uniqueness of solutions of ultrahyperbolic partial differential equations,	184
PA, CHENKUO. A new definition of the Godeaux sequence of quadrics, . . .	117
RAJAGOPAL, C. T. On Riesz summability and summability by Dirichlet's series,	371
RAJAGOPAL, C. T. On Riesz summability and summability by Dirichlet's series,	851
ROBINSON, G. DE B. On the representations of the symmetric group (second paper),	28
SAMUEL, P. On conformal correspondence of surfaces and manifolds, . . .	4
SCHUR, ISSAI. Identities in the theory of power series,	
SCOTT, W. T. and H. S. WALL. On the convergence and divergence of continued fractions,	55
SNAPPER, ERNST. Polynomial matrices in one variable, differential equations and module theory,	29
SNAPPER, ERNST. Polynomial matrices in several variables,	62

	PAGE
STEWART, B. M. Left associated matrices with elements in an algebraic domain,	562
STRODT, WALTER. Principal solutions of difference equations, . . .	717
THOMAS, JOSEPH MILLER. Eliminants,	592
WALL, H. S. See Scott and Wall, page 551.	
WINTNER, AUREL. (L^2)-connections between the potential and kinetic energies of linear systems,	5
WINTNER, AUREL. On the Laplace-Fourier transcendents occurring in mathematical physics,	87
WINTNER, AUREL. On the Tauberian nature of Ikehara's theorem, . . .	99
WINTNER, AUREL. Asymptotic integrations of the adiabatic oscillator, . . .	251
WINTNER, AUREL. The sum formula of Euler-Maclaurin and the inversions of Fourier and Möbius,	685
WINTNER, AUREL. On Töpler's wave analysis,	758
WINTNER, AUREL. On Riemann's reduction of Dirichlet series to power series,	769
WINTNER, AUREL. Vortices and nodes,	815
WINTNER, AUREL. See Hartman and Wintner, page 853.	
WU, GEORGE. Study of a surface by means of certain associate ruled surfaces in affine space,	801
ZYGMUND, A. On the summability of multiple Fourier series, . . .	836

	PAGE
FERMAN, PHILIP. Toulner's theorem and absolute constants,	539
HARTMAN, PHILIP and ARTHUR WILKINS. On Möbius' inversion,	855
HAYKANA, A. K. A note on Laplace transforms of functions whose spectra are confined to a given set,	219
HSIUNG, CHUAN-CHIH. A study on the theory of conjugate nets,	319
HSIUNG, CHUAN-CHIH. Projective theory of surfaces and conjugate nets in four-dimensional space,	667
JACOBSON, N. A note on division rings,	27
KAGENO, T. N. Desargues' and Pappus' graphs and their groups,	859
KAGENO, T. N. Commutators,	372
KAPLAN-KY, JEROME. Topological rings,	152
KOSNIER, EDWARD and JOHN DE CICCIO. Extensions of harmonic transformations,	575
LEITCH, J. I. Simple links and fixed sets under continuous mappings,	345
LEE, HWA-CHUNG. On skew-metric spaces and function groups,	790
LEWIS, F. C. Polynomial least square approximations,	273
MANUEL, JULIAN D. Unilateral variations with variable end-points,	121
McCoy, NATHAN. See L. Howland McCoy, page 46.	
MERVALE, JAMES CLAYTON. Inequalities of higher degree in one or two unknowns,	35
MILLER, EDWARD. The finite modular function and a class of analytic functions first considered by Hurwitz,	10
OWENS, D. G. Uniqueness of solutions of ultrahyperbolic partial differential equations,	184
PA, CHEN-CHAO. A new definition of the Gödelian sequence of quadrants,	777
PERJACOFF, C. T. On Riesz summability and summability by Dini's method,	373
PERJACOFF, C. T. On Riesz summability and summability by Dini's method,	85
ROBINSON, GEORGE B. On the representations of the symmetric group (second paper),	
SAMUEL, P. On conformal correspondence of surfaces and manifolds,	
SCHUB, LOUIS. Identities in the theory of power series,	
SEITZ, W. TH. and H. S. WARD. On the convergence and acceleration of continued fractions,	53
SNAPPER, ERNST. Polynomial matrices in one variable differential equations and module theory,	200
SNAPPER, ERNST. Polynomial matrices in several variables,	627

36
CUT. Hc 1635-61. P. 1411.
(AMERICAN (61))

JOURNAL OF MATHEMATICS

FOUNDED BY THE JOHNS HOPKINS UNIVERSITY

EDITED BY

R. BRAUER
UNIVERSITY OF TORONTO

F. D. MURNAGHAN
THE JOHNS HOPKINS UNIVERSITY

L. M. GRAVES
UNIVERSITY OF CHICAGO

H. WHITNEY
HARVARD UNIVERSITY

THE JOH

VERSITY

WITH THE COOPERATION OF

S. BOCHNER
D. MONTGOMERY
C. L. SIEGEL
D. C. SPENCER
M. A. ZORN

BIRKHOFF
COURANT
P. HARTMAN
J. L. VANDERSLICE
A. ZYGMUND

L. BERS
J. L. DOOB
N. JACOBSON
D. C. LEWIS
A. D. WALLACE

PUBLISHED UNDER THE JOINT AUSPICES OF

THE JOHNS HOPKINS UNIVERSITY

AND

THE AMERICAN MATHEMATICAL SOCIETY

(VOLUME LXIX)
(1947)

THE JOHNS HOPKINS PRESS

BALTIMORE 18, MARYLAND.

U. S. A.

916908

PRINTED IN THE UNITED STATES OF AMERICA
BY J. H. FURST COMPANY, BALTIMORE, MARYLAND

THE FREE PRODUCT OF GROUPS.*

By EMIL ARTIN.

The traditional way of introducing the free product of groups consists in defining carefully the product of elements and proving the associative law. We shall use here, on the contrary, a method where the associative law is obvious and all the difficulties arise from the discussion of equality.

Let Σ be a set of groups Γ . Without loss of generality, we assume these groups to be disjoint. The elements of all these groups will be denoted by small letters a, b, c, \dots whereas the capitals A, B, \dots stand for symbolic products

$$(1) \quad A = a_1 a_2 \cdots a_n.$$

The elements a_i may or may not belong to the same Γ . The number n of terms is called the length of A . Two such products are equal if and only if they consist of the same factors in precisely the same order; in other words, no simplification is allowed in (1), not even if two adjoining elements belong to the same Γ or if one of the factors is a unit. Multiplication AB is defined in the obvious way by writing first the factors of A and then those of B . The associative law is therefore trivial. It is convenient to introduce also the "empty" product E which acts as unit element. Nevertheless we do not get a group since $A \neq E$ does not have an inverse element. In spite of this we use as pure notation A^{-1} for the product $a_n^{-1} a_{n-1}^{-1} \cdots a_1^{-1}$.

It might happen that all the a_i in (1) are from the same group Γ and that the product $a_1 a_2 \cdots a_n$ computed in Γ has the value 1. We then call A an elementary expression.

Let us now define a certain set Δ of formal products. It contains those products that we want later to equate to E and is defined by induction:

1) E is in Δ .

2) If A has a positive length it belongs to Δ if and only if it can be written in the form

$$(2) \quad A = A_0 a_1 A_1 a_2 \cdots A_{n-1} a_n A_n$$

* Received May 10, 1946.

916908

PRINTED IN THE UNITED STATES OF AMERICA
BY J. H. FURST COMPANY, BALTIMORE, MARYLAND

where all A_i are elements of Δ (and have smaller length, of course) and where the a_i are such that $a_1 a_2 \cdots a_n$ is an elementary expression.

We prove now a few lemmas about our set Δ .

1) $A \in \Delta$ implies $A^{-1} \in \Delta$; $A \in \Delta$ and $B \in \Delta$ implies $AB \in \Delta$. Proof by induction. Both parts are true for $A = E$. Let A be of the form (2). Every A_i is shorter than A so that the theorem may be assumed for A_i . This proves directly the first half. For the second half we can assume $A_n B \in \Delta$. Now AB has again the form (2) with $A_n B$ instead of A_n .

2) If $A \in \Delta$ and $A \neq E$ we can always assume that A is of the form (2) with $A_0 = E$ (or also with $A_n = E$ if we prefer). Proof by induction. Assume A of form (2) and denote by B the whole product (2) but without A_0 . Then $B \in \Delta$ and $A = A_0 B$. If $A_0 = E$ we have nothing to prove; at any rate A_0 is shorter than A so that we may assume $A_0 = b_1 B_1 b_2 B_2 \cdots B_n$. Now $A = b_1 B_1 b_2 B_2 \cdots b_n (B_n B)$.

3) If the first and the last term of an $A \in \Delta$ belong to the same Γ , we can even assume in (2) that $A_0 = A_n = E$.

Proof. We can assume $A_0 = E$; now $A_n \in \Delta$ so that we can write A_n in the form

$$A_n = B_0 b_1 B_1 b_2 \cdots b_m.$$

This leads to

$$A = a_1 A_1 a_2 A_2 \cdots a_n B_0 b_1 B_1 b_2 \cdots B_{m-1} b_m.$$

Since $a_1 a_2 \cdots a_n b_1 \cdots b_m$ is elementary, A is of the required form.

4) aAa^{-1} and A belong to Δ or do not at the same time. *Proof.* a) Let $A \in \Delta$. aAa^{-1} is elementary so $aAa^{-1} \in \Delta$. b) Let $aAa^{-1} \in \Delta$. According to a previous lemma it is of the form

$$aAa^{-1} = aA_1 a_2 A_2 \cdots a_{n-1} A_{n-1} a^{-1}$$

so that we have

$$A = A_1 a_2 A_2 \cdots a_{n-1} A_{n-1} \text{ and } a_2 a_3 \cdots a_{n-1} \text{ is elementary.}$$

Repeated application proves:

5) BAB^{-1} and A belong to Δ or do not at the same time.

INDEX

ALBERTI, EUGENIO. Sul calcolo delle soluzioni dei problemi al contorno per le equazioni lineari del secondo ordine di tipo ellittico.	17
ALLEN, JOHN. The free product of groups.	4
BAUM, MAXIMILIAN. The double chain condition in cyclic operator groups.	51
BARTO, RICHARD. Projectivities of finite projective planes.	65
BIRNBAUM, ZOSHA. On some triangular summability methods.	12
BIRNBAUM, ZOSHA. Free loops and nets and their generalizations.	1
BIRNBAUM, ZOSHA. Six-rings in minimal five-color maps.	391
BIRNBAUM, ZOSHA. Sierpinski sets.	46
BIRNBAUM, ZOSHA. On the zeta-functions of algebraic number fields.	2
BIRNBAUM, ZOSHA. Applications of induced characters.	70
BIRNBAUM, ZOSHA. NEAL H. MCCOY. Radicals and subdirect sums.	3
BIRNBAUM, ZOSHA. ELLLEN F. BUCK. A note on finitely-additive measures.	413
BIRNBAUM, ZOSHA. See ELLEN F. BUCK and BIRNBAUM, ZOSHA, page 413.	
BIRNBAUM, ZOSHA. The isoperimetric problem in the Minkowski plane.	843
BIRNBAUM, ZOSHA. Note on some problems in number theory.	18
BIRNBAUM, ZOSHA. R. Generic schemes.	323
BIRNBAUM, ZOSHA. Union-preserving transformations of higher order arrangements.	104
BIRNBAUM, ZOSHA. See KASPER and DE CICCIO, page 575.	
BIRNBAUM, ZOSHA. Mapping theorems for non-compact spaces.	20
BIRNBAUM, ZOSHA. Analyticity in Hilbert space and self-adjoint transformations.	60
BIRNBAUM, ZOSHA. On convergence of the product of basic sets of polynomials.	533
BIRNBAUM, ZOSHA. Sur une suite de quadriques associée à une confluence H .	490
BIRNBAUM, ZOSHA. On congruences and conjugate nets.	70
BIRNBAUM, ZOSHA. On the ergodic theorems.	103
BIRNBAUM, ZOSHA. Systems of total differential equations and Liouville's problem on conformal mappings.	107

6) $AB \in \Delta$ implies $BA \in \Delta$. *Proof.* $AB \in \Delta$ implies $AB \cdot AA^{-1} \in \Delta$ since $AA^{-1} = AEA^{-1}$ belongs to Δ . From 5) we get now $BA \in \Delta$.

DEFINITION. A and B are called congruent, $A \equiv B$, if and only if $AB^{-1} \in \Delta$ ($B^{-1}A \in \Delta$ would mean the same).

We see at once:

- a) $A \equiv A$ since $AA^{-1} \in \Delta$.
- b) $A \equiv B$ implies $B \equiv A$. Indeed $AB^{-1} \in \Delta$ implies $BA^{-1} \in \Delta$.
- c) $A \equiv B$ and $B \equiv C$ implies $A \equiv C$, because $AB^{-1} \in \Delta$ and $BC^{-1} \in \Delta$ imply $BC^{-1}AB^{-1} \in \Delta$ and therefore $C^{-1}A \in \Delta$.
- d) $A \equiv B$ and $C \equiv D$ imply $AC \equiv BD$. Indeed $B^{-1}A \in \Delta$ and $CD^{-1} \in \Delta$ imply $B^{-1}ACD^{-1} \in \Delta$ and therefore $ACD^{-1}B^{-1} \in \Delta$.
- e) $AA^{-1} \equiv E$.

THEOREM 1. *Our formal products form a group if we use the congruence instead of equality. The elements of Δ are precisely those congruent to E .*

Let us now consider the elements $A = a$ of length 1. Two of them are congruent, $a \equiv b$, if and only if ab^{-1} is in Δ . Either ab^{-1} has to be elementary, which means $a = b$, or both factors have to be in Δ and therefore to be the unit elements of their group. Our congruence equates, therefore, only the unit elements among our elements a . We therefore denote from now on all unit elements by 1. Let next a and b belong to the same group Γ and let c be their product in Γ . The formal product ab is then congruent to c because abc^{-1} is elementary.

Our group contains, therefore, each of the groups Γ as a subgroup, and different Γ 's have only the unit in common.

In the sense of congruence we are now in the position to simplify a symbolic product A by uniting two adjoining factors as soon as they belong to the same group and by dropping every unit element. Since these operations reduce the length of A we finally arrive either at E or at an expression where no factor is 1 and no two adjoining terms are of the same group. We call such an expression normal.

	PAGE
HARTMAN, PHILIP. Tauber's theorem and absolute constants, . . .	599
HARTMAN, PHILIP and AUREL WINTNER. On Möbius' inversion, . . .	853
HAVILAND, E. K. A note on Laplace transforms of functions whose spectra are confined to a given set, . . .	279
HSIUNG, CHUAN-CHIH. A study on the theory of conjugate nets, . . .	379
HSIUNG, CHUAN-CHIH. Projective theory of surfaces and conjugate nets in four-dimensional space, . . .	607
JACOBSON, N. A note on division rings, . . .	27
KAGNO, I. N. Desargues' and Pappus' graphs and their groups, . . .	859
KAGNO, I. N. Corrections, . . .	852
KAPLANSKY, IRVING. Topological rings, . . .	153
KASNER, EDWARD and JOHN DE CICCO. Extensions of harmonic transformations, . . .	575
KELLEY, J. L. Simple links and fixed sets under continuous mappings, . . .	348
LEE, HWA-CHUNG. On skew-metric spaces and function groups, . . .	790
LEWIS, D. C. Polynomial least square approximations, . . .	273
MANCILL, JULIAN D. Unilateral variations with variable end-points, . . .	121
MCCOY, NEAL H. See Brown and McCoy, page 46.	
MESERVE, BRUCE ELWYN. Inequalities of higher degree in one unknown, . . .	357
NEILARI, ZEEV. The elliptic modular function and a class of analytic functions first considered by Hurwitz, . . .	70
OWENS, O. G. Uniqueness of solutions of ultrahyperbolic partial differential equations, . . .	184
PA, CHENKUO. A new definition of the Godeaux sequence of quadrics, . . .	117
RAJAGOPAL, C. T. On Riesz summability and summability by Dirichlet's series, . . .	371
RAJAGOPAL, C. T. On Riesz summability and summability by Dirichlet's series, . . .	851
ROBINSON, G. DE B. On the representations of the symmetric group (second paper), . . .	27
SAMUEL, P. On conformal correspondence of surfaces and manifolds, . . .	4
SCHUR, ISSAI. Identities in the theory of power series, . . .	
SCOTT, W. T. and H. S. WALL. On the convergence and divergence of continued fractions, . . .	55
SNAPPER, ERNST. Polynomial matrices in one variable, differential equations and module theory, . . .	291
SNAPPER, ERNST. Polynomial matrices in several variables, . . .	62

THEOREM 2. *Every A is congruent to a normal form.*

LEMMA. *A normal form $A = a_1 a_2 \cdots a_n$ of length > 0 cannot be $\equiv 1$.*

Proof by induction. Assume it true for shorter expressions. If we write A in the form (2) all the A_i are shorter than A and also normal (as parts of A). So they must be E and only $A = a_1 a_2 \cdots a_n$ remains and has to be elementary. But that is impossible.

If we ask now when two normal expression can be congruent, $a_1 a_2 \cdots a_n \equiv b_1 b_2 \cdots b_m$, we have to consider whether $a_1 a_2 \cdots a_n b_m^{-1} b_{m-1}^{-1} \cdots b_1^{-1}$ can be $\equiv 1$. Since this product must never become a normal expression before being reduced to E we easily deduce:

THEOREM 3. *Two normal expressions are congruent if and only if they are identical.*

PRINCETON UNIVERSITY,
PRINCETON, N. J.

INDEX.

SHIMMURA, B. M. Unit associated matrices with elements in integral domain,	121
SHIMURA, G. Principal solutions of difference equations,	122
SHIMURA, G. and MILLER. Elimination,	123
SHIMURA, H. See Scott and Wall, page 551.	
SHIMURA, A. (5 ²)-connections between the potential and kinetic energies of linear systems,	124
SHIMURA, G. On the Laplace-Fourier transcendents occurring in mathematical physics,	125
SHIMURA, A. On the Tauberian nature of Ikibara's theorem,	126
SHIMURA, A. Asymptotic integrations of the adiabatic oscillator,	127
SHIMURA, A. The sum formula of Euler-Maclaurin and the inversions of Fourier and Möbius,	128
SHIMURA, A. On Töpler's wave analysis,	129
SHIMURA, A. On Riemann's reduction of Dirichlet series to power series,	130
SHIMURA, A. Vortices and nodes,	131
SHIMURA, A. See Hartman and Wintner, page 853	
SHIMURA, G. Study of a surface by means of certain associated surfaces in affine space,	132
SHIMURA, A. On the summability of multiple Fourier series,	133

(L^2) -CONNECTIONS BETWEEN THE POTENTIAL AND KINETIC ENERGIES OF LINEAR SYSTEMS.*

By AUREL WINTNER.

1. Let primes denote differentiations with respect to a real variable, t , the range of which will be the half-line $t \geq 0$. The functions to be considered, $f(t)$, $g(t)$, $x(t)$, will be assumed to be defined and continuous on this half-line. For the sake of simplicity, it will be assumed from the beginning that all these functions are scalars. In addition, *all* functions will be regarded as real-valued.

Correspondingly, $x(t)$ will be said to be of class (L^2) if

$$\int_0^{\infty} x^2(t) dt < \infty.$$

This is neither necessary nor sufficient in order that $x(t)$ be of class $(L) = (L^1)$:

$$\int_0^{\infty} |x(t)| dt < \infty.$$

In fact, one of the principal difficulties connected with the oscillation problems to be considered is that

$$(1) \quad x(t) \rightarrow 0 \text{ as } t \rightarrow \infty$$

cannot be assumed.

The results will deal with the (L^2) -solutions, $x(t)$, of linear differential equations

$$(2) \quad x'' + f(t)x = 0$$

and of the corresponding inhomogeneous equations

$$(3) \quad x'' + f(t)x = g(t).$$

The issue will be the (L^2) -character of the derivative, $x'(t)$. Such problems suggest themselves by the usual formal treatment of the characteristic functions, $x(t)$, of a Schrödinger equation (where $t = r$). For, on the one hand, neither

* Received September 25, 1946.

$$\int_0^{\infty} x'^2(t) dt < \infty$$

nor

$$x'(t) \rightarrow 0 \text{ as } t \rightarrow \infty$$

should be assumed of a characteristic function and, on the other hand, the partial integrations (on which, for instance, the deduction of the relation of indeterminacy depends) cannot be made legitimate otherwise.

In order to avoid this difficulty, I gave (iii) below, under the restriction (4), in a lecture many years ago. I noticed only recently that what is actually needed when "locating" continuous spectra, namely, (iv) below, can be concluded from (iii).

2. If (2) is interpreted as a dynamical system, then $\frac{1}{2}x'^2$ is the kinetic, and $\frac{1}{2}fx^2$ the potential, energy of a solution $x = x(t)$. Their sum is not a constant, since the Lagrangian equation, (2), contains t explicitly. To say that $x'(t)$ is of class (L^2) means that the sum total of the *density* of the kinetic energy remains finite as $t \rightarrow \infty$. The corresponding remark for $x(t)$ and the potential energy, $\frac{1}{2}fx^2$, holds only if $|f|^{\frac{1}{2}}$, the absolute frequency of (2), does not come "too close" to 0 and ∞ , as $t \rightarrow \infty$.

This suggests that, under an appropriate restriction of the behavior of $f(t)$ as $t \rightarrow \infty$, the (L^2) -character of $x'(t)$ can be deduced from that of $x(t)$, if use is made of a form of those "virial" considerations which, in the latter part of the fourth of his *Vorlesungen über Dynamik*, Jacobi has applied to the (non-linear, but conservative) case of a solar system. What can actually be proved in this manner is the second of the assertions of the following theorem:

(i) If, as $t \rightarrow \infty$,

$$(4) \quad f(t) = O(1),$$

then a solution, $x(t)$, of (1) cannot be of class (L^2) unless $x'(t)$ is also of class (L^2) . The same is true if (4) is relaxed to

$$(5) \quad f(t) = O_R(1).$$

It is not in general true if (4) is omitted entirely; in fact, even $f(t) > 0$, hence

$$(5 \text{ bis}) \quad f(t) = O_L(1),$$

is insufficient.

In (5) and (5 bis), the subscripts refer to boundedness from above ("right") and from below ("left"), respectively.

The insufficiency of the restriction (5 bis) is proved by the example

$$(5^*) \quad f(t) = (e^t + \tfrac{1}{2})(e^t - \tfrac{1}{2}) > 0, \quad (0 \leq t < \infty).$$

For, in this case, (2) is readily verified to have the solution $x(t) = e^{-\frac{1}{2}t} \cos e^t$. Clearly, this solution is of class (L^2) , and (5 bis) is satisfied. Nevertheless, $x'(t)$ is not of class (L^2) . In fact, $x'(t)$ is the sum of two terms, one of which is $O(e^{-\frac{1}{2}t})$, hence a function of class (L^2) , whereas the other term is $-e^{\frac{1}{2}t} \sin e^t$, a function which, in view of

$$\int_0^\infty (-e^{\frac{1}{2}t} \sin e^t)^2 dt = \int_1^\infty \sin^2 s \, ds = \infty, \quad (s = e^t),$$

fails to be of class (L^2) .

3. The verification of the preceding solution makes it clear that, if $f(t)$ is given by (5*), both

$$x(t) = e^{-\frac{1}{2}t} \cos e^t \text{ and } x(t) = e^{-\frac{1}{2}t} \sin e^t$$

are solutions of (2). Since both of these solutions are of class (L^2) , the example (5*) proves the second of the assertions of the following theorem:

(ii) *If $f(t)$ is non-negative and satisfies, uniformly for all large t , a Lipschitz condition,*

$$|f(t_1) - f(t_2)| \leq \text{const.} \cdot |t_1 - t_2|,$$

then (2) cannot have two linearly independent solutions of class (L^2) . The same is not in general true if the Lipschitz condition (which, incidentally, allows unbounded functions $f(t)$, such as $t^{\frac{1}{2}}$ or t , and even $t \sin^2 1/t$) is omitted, that is, if only (5 bis) or $f(t) \geq 0$ is assumed.

On the other hand, the Lipschitz condition is superfluous if (5 bis) is replaced by the assumption, (5), of (i). More generally, (2) cannot have two linearly independent solutions of class (L^2) if $f(t)$ has the property that no solution, $x(t)$, of (2) is of class (L^2) unless $x'(t)$ also is of class (L^2) .

The proof of the first assertion of (ii) proceeds as follows:

It is easily verified (for instance, by a partial integration of the relevant Stieltjes integral) that

$$d\{x'^2(t) + f(t)x^2(t)\} = x^2(t)df(t)$$

is an identity in t along any solution, $x(t)$, of (2) (the notation $df(t)$ is justified in itself, since $f(t)$, being subject to a Lipschitz condition, is absolutely continuous). Clearly, the last two formula lines imply the inequality

$$\int_0^\infty |d\{x'^2(t) + f(t)x^2(t)\}| \leq \text{const.} \int_0^\infty x^2(t) dt.$$

Hence, if $x(t)$ is of class (L^2) , the integral on the left of this inequality is convergent. This means that the function $x'^2(t) + f(t)x^2(t)$ is of bounded variation. In particular, it is a bounded function. It follows therefore from the first assumption, $f(t) \geq 0$ (which has not been used thus far), that $x'^2(t)$ is bounded. In other words, a solution, $x(t)$, cannot be of class (L^2) unless $x'(t) = O(1)$.

Consequently, if $x = x(t)$ and $y = y(t)$ are two solutions of class (L^2) , then $x'(t) = O(1)$ and $y'(t) = O(1)$, and so the Wronskian, $x'(t)y(t) - y'(t)x(t)$, is of class (L^2) . But the Wronskian of any two solutions of (2) is independent of t (Abel). Since a function which is independent of t is of class (L^2) on the half-line $0 \leq t < \infty$ only if it vanishes identically, it follows that the Wronskian of $x(t)$ and $y(t)$ is 0. Accordingly, $x(t)$ and $y(t)$ are linearly dependent. This proves the first assertion of (ii).

The fourth assertion of (ii) (which, in view of (i), implies the third), can be concluded by a slight modification of the last argument. For, if $x = x(t)$ and $y = y(t)$ are of class (L^2) , then, by the present assumption, the same is true of $x'(t)$ and $y'(t)$. But the product of two functions of class (L^2) is absolutely integrable (Schwarz), hence integrable. Consequently, the Wronskian, although a constant, is integrable on the half-line $0 \leq t < \infty$ and must, therefore, vanish. Since this means that $x(t)$ and $y(t)$ are linearly dependent, the last assertion of (ii) follows.

4. The third assertion of (i) has been proved by the example (5*). The second assertion of (i), which remains to be proved, implies the following

COROLLARY. *If $f(t)$ is restricted by (5), then a solution, $x(t)$, of (2) cannot be of class (L^2) unless $x(t)$ satisfies (1).*

First, (i) and the assumptions of the Corollary imply that both functions $x(t)$, $x'(t)$ are of class (L^2) . Hence, the product $x(t)x'(t)$ is absolutely integrable, and so its integral over the half-line $0 \leq t < \infty$ is convergent. But, since $xx' = (\frac{1}{2}x^2)'$, the convergence of this improper integral means that

the square of $x(t)$ tends to a finite limit as $t \rightarrow \infty$. Finally, this limit cannot be distinct from 0, since $x(t)$ is of class (L^2).

5. The remaining part of (i) is a particular case, $g \equiv 0$, of the following theorem:

(iii) *If $f(t)$ is restricted by (4), or merely by (5), and if $g(t)$ is of class (L^2), then a solution, $x(t)$, of (3) cannot be of class (L^2) unless $x'(t)$ is also of class (L^2).*

Only the "external force," $g(t)$, and not the main coefficient function, $f(t)$, of (2) as well, is required to be of class (L^2).

Since x and g are supposed to be of class (L^2), the product xg is of class (L). Hence, (3) shows that the sum $xx'' + fx^2$ is of class (L). In order to avoid a secondary complication which arises when (4) is generalized to (5), suppose first (4). Then, since x is of class (L^2), hence x^2 of class (L), the product fx^2 is of class (L). But the same is true of the sum $xx'' + fx^2$. Consequently, xx'' is of class (L).

However, xx'' is identical with $(xx')' - x'^2$. On the other hand, the assertion of (iii), according to which x' is of class (L^2), means that x'^2 is of class (L). Hence, if the assertion were not true, then, since $(xx')' - x'^2$ is of class (L), and since $x'^2 \geq 0$, it would follow that the indefinite integral of $(xx')'$ must tend to ∞ as $t \rightarrow \infty$. But this leads to a contradiction.

In fact, the indefinite integral just mentioned is $xx' + \text{const.}$ Since $xx' = (\frac{1}{2}x^2)'$, it follows that what is supposed to tend to ∞ is $(\frac{1}{2}x^2)'$. But a quadrature shows that $(x^2)' \rightarrow \infty$ implies $x^2 \rightarrow \infty$, as $t \rightarrow \infty$. However, $x^2 \rightarrow \infty$ contradicts the assumption that $x(t)$ is of class (L^2). This proves (iii) in the particular case, (4), of (5).

In order to relax (4), put $f(t) = p(t) + n(t)$, where $p = \max(f, 0)$. Then both functions p , $-n$ are non-negative (and continuous, since f is). Furthermore, since the assumption (5) means that p is bounded, and since x is of class (L^2), the product px^2 is of class (L). But $xx'' + fx^2$ was seen to be of class (L) (without any assumption on f). It follows therefore from $fx^2 = px^2 + nx^2$ that the sum $xx'' + nx^2$ is of class (L).

In particular, the indefinite integral of this sum tends to a finite limit as $t \rightarrow \infty$. If the indefinite integral of the second term, nx^2 , of the sum does not tend to a finite limit, then, since $nx^2 \leq 0$, it must tend to $-\infty$, and so the indefinite integral of the first term, xx'' , must tend to ∞ . Since xx'' , being identical with $(xx')' - x'^2$, is not less than $(xx')'$, this leads to the same contradiction as before.

Consequently, the indefinite integral of nx^2 cannot tend to $-\infty$. Since $nx^2 \leq 0$, and since $xx'' + nx^2$ is of class (L) , it follows that xx'' is of class (L) . The balance of the proof is the same as before.

6. It will be convenient to use the following abbreviated manner of speaking:

DEFINITION. Let $f(t)$ be called of class $(*)$ if every solution of the corresponding homogeneous differential equation, (2), satisfies

$$(6) \quad x(t) = O(1) \text{ and } x'(t) = O(1) \quad (t \rightarrow \infty).$$

There are various explicit criteria sufficient in order that $f(t)$ be of class $(*)$. For instance, it is known that $f(t)$ is sure to be of class $(*)$ if there exists a positive constant, ω^2 , satisfying either

$$\int_0^\infty |f(t) - \omega^2| dt < \infty$$

(where $0 < \limsup_{t \rightarrow \infty} |f(t) - \omega^2| \leq \infty$ is allowed) or

$$\int_0^\infty |df(t)| < \infty \text{ and } f(t) \rightarrow \omega^2 \text{ as } t \rightarrow \infty$$

(these two criteria are independent, even if $f(t) \rightarrow \omega^2$ is assumed in the first). However, in the following theorem, the property $(*)$ of $f(t)$ can be thought of as assured by some criterion.

(iv) Let $f(t)$ be a function which is of class $(*)$ and satisfies (4) or, more generally, (5), and let $g(t)$ be a fixed function of class (L^2) . Then the inhomogeneous equation (3), belonging to the given $g(t)$, possesses a solution of class (L^2) if and only if there belongs to every solution, $x(t)$, of the homogenous equation, (2), a (unique) constant, $c = c_{x(t)}$, for which the function

$$(8) \quad c - \int_0^t x(u)g(u)du$$

becomes a function of class (L^2) .

The existence of such a constant c is required for every solution of (2), rather than for only those solutions of (2) which are of class (L^2) .

The point in (iv) is that, in contrast to what is usually assumed in the theory of spectra, the (L^2) -requirement is now placed only on a solution of

the homogeneous equation, rather than on such a solution *and* on its derivative. The rôle of (iii) is precisely this reduction.

7. The proof of (iv) proceeds as follows:

Let $x(t)$ denote a *fixed* solution of (2). If $x(t) \equiv 0$, then the constant the existence of which is claimed in (iv) clearly is $c = 0$. It can therefore be assumed that $x(t) \not\equiv 0$. Then (2) has a solution, say $y(t)$, which is linearly independent of $x(t)$. This means that the determinant of the Wronskian matrix,

$$(9) \quad W(t) = \begin{pmatrix} y(t) & x(t) \\ y'(t) & x'(t) \end{pmatrix},$$

does not vanish. Since $\det W(t)$ is independent of t (Abel), and since $y(t)$ can be multiplied by any non-vanishing constant, $y(t)$ can be chosen as follows: $\det W(t) \equiv 1$. Then the inverse of the matrix (9) is

$$(10) \quad W^{-1}(t) = \begin{pmatrix} x'(t) & -x(t) \\ -y'(t) & y(t) \end{pmatrix}.$$

Since $x(t)$ and $y(t)$ are solutions of the homogeneous equation (2), it is easily verified from (10) that, if $\phi(t)$ denotes the vector

$$(11) \quad \phi(t) = \int_0^t W^{-1}(u) \begin{pmatrix} 0 \\ g(u) \end{pmatrix} du,$$

then the first component of the vector $W(t)\phi(t)$ is a solution, $x = x_0(t)$, and the second component of (11) the derivative, $x_0'(t)$, of this solution, of the inhomogeneous equation, (3). (Needless to say, the differentiations occurring in this verification are equivalent to the proof of the rule for the variation of constants.) On the other hand, since either column of (9) is a vector the first component of which is a solution, and the second the derivative of this solution, of (2), the general solution of (2) results if an arbitrary constant vector is transformed by the matrix $W(t)$. Hence, if a and b denote the components of the arbitrary constant vector, then the first component of the vector

$$(12) \quad W(t)\phi(t) + W(t) \begin{pmatrix} a \\ b \end{pmatrix}$$

represents an arbitrary solution, and the second component the derivative of this solution, of (3).

By assumption, $f(t)$ is class (*). Since the definition of this class requires (6) not only for the given $x = x(t)$ but for $x = y(t)$ as well, it follows from (9) and (10) that all elements of both matrices $W(t)$, $W^{-1}(t)$ are $O(1)$ as $t \rightarrow \infty$. Consequently, both components of the vector (12) are of class (L^2) (for fixed a, b) if and only if both components of the vector

$$\phi(t) + \begin{pmatrix} a \\ b \end{pmatrix}$$

are of class (L^2) for the same values of the scalar constants a, b . For, on the one hand, these two vectors are transformed into one another by the matrices $W(t)$, $W^{-1}(t)$ and, on the other hand, the sum of two functions of class (L^2) , as well as the product of two functions one of which is of class (L^2) and the other $O(1)$, is of class (L^2) .

Accordingly, a solution of (3) is of class (L^2) along with its derivative if and only if the pair of integration constants a, b which belong to it by virtue of (12) is such as to render both components of the vector represented by the last formula line functions of class (L^2) . However, since the assumptions made in (iii) with regard to $f(t)$ and $g(t)$ are assumed in (iv), the restriction just italicized, that concerning the derivative of a solution of (3), is made superfluous by (iii). On the other hand, (11) and (12) show that the vector in question is

$$\int_0^t \begin{pmatrix} x'(s) & -x(s) \\ -y'(s) & y(s) \end{pmatrix} \begin{pmatrix} 0 \\ g(s) \end{pmatrix} ds + \begin{pmatrix} a \\ b \end{pmatrix},$$

which means that the first component of the vector is

$$a - \int_0^t x(s)g(s)ds$$

and the second the negative of

$$-b - \int_0^t y(s)g(s)ds.$$

Since both of these expressions are of the form (8), the proof of (iv) is complete.

8. Under the first assumption of (i), the content of (i) can be amplified as follows:

(i*) *If $f(t)$ satisfies (4) and if a solution, $x(t)$, of (2) is of class (L^2) , then both this solution and its derivative must tend to 0, whereas no solution linearly independent of this solution can remain bounded, as $t \rightarrow \infty$.*

The first of these three assertions, viz., (1), has been deduced above as a corollary of (i). In fact, only (5), rather than (4), was assumed in this deduction. The following proof of the second assertion, viz., of

$$(1^*) \quad x'(t) \rightarrow 0 \text{ as } t \rightarrow \infty,$$

leaves it undecided whether or not (4) can again be relaxed to (5).

First, (i) implies that the product xx' is of class (L) . It follows therefore from (4) that the same is true of f times xx' . In view of (2), this means that $x'x''$ is of class (L) . Consequently, the indefinite integral of $x'x''$ must tend to a finite limit as $t \rightarrow \infty$. Since this indefinite integral is $\frac{1}{2}x'^2$, it follows that, if the assertion of the last formula line did not hold, $x'(t)$ could not be of class (L^2) .

In order to prove the remaining assertion of the last italicized theorem, let $x = y(t)$ be a solution of (2) satisfying $y(t) = O(1)$. The assertion is that $y(x)$ and $x(t)$ must be linearly dependent.

What this assertion claims is the (identical) vanishing of the Wronskian, $xy' - yx'$, the latter being independent of t for any two solutions of (2). Hence, it is sufficient to show that $xy' - yx'$ tends to 0 as $t \rightarrow \infty$. But both x and x' tend to 0 as $t \rightarrow \infty$. Consequently, it is sufficient to ascertain that both y and y' are $O(1)$. But $y(t) = O(1)$ is assumed for the solution $x = y(t)$ of (2), and so (4) and (2) imply that $y''(t) = O(1)$. This completes the proof, since $y'(t) = O(1)$ is a necessity for every function $y(t)$ having a second derivative and satisfying both $y(t) = O(1)$ and $y''(t) = O(1)$, as $t \rightarrow \infty$ (Hadamard).

IDENTITIES IN THE THEORY OF POWER SERIES.*

By ISSAI SCHUR.¹

1. Introduction. Let

$$(1) \quad g(x) = 1 + a_1x + a_2x^2 + \cdots + a_nx^n + \cdots$$

be a power series which converges in a certain neighborhood of $x = 0$. We write

$$(2) \quad [g(x)]^n = \sum_{\rho=0}^{\infty} a_{n\rho}x^\rho, \quad (a_{n0} = 1); \quad n = 0, \pm 1, \pm 2, \cdots,$$

and introduce the infinite set of series

$$(3) \quad \phi_n(x) = \sum_{\nu=\sigma}^{\infty} a_{n-\nu, \nu}x^\nu; \quad n = 0, \pm 1, \pm 2, \cdots.$$

We wish to prove the following theorems:

I. The ratio

$$(4) \quad \frac{\phi_{n+1}(x)}{\phi_n(x)} = \psi(x)$$

is independent of n .

II. If we put $f(x) = xg(x)$, then

$$(5) \quad \psi[f(x)] = g(x).$$

III. For every $n = 0, \pm 1, \pm 2, \cdots$

$$(6) \quad f'(x)\phi_n[f(x)] = [g(x)]^{n+1}.$$

In order to establish the truth of these propositions, it suffices evidently to prove Theorem III. For, dividing two identities (6) for $n+1$ and n we get the identity (5) with $\psi(x) = \frac{\phi_{n+1}(x)}{\phi_n(x)}$; this proves further the inde-

* Received June 10, 1946.

¹ Died January 10th, 1941, Tel-Aviv, Palestine. This is the third posthumous paper of the deceased. The preceding two papers were: 1. "On Faber polynomials," *American Journal of Mathematics*, vol. 67 (1945), pp. 33-41. 2. "Ein Satz über quadratische Formen," *American Journal of Mathematics*, vol. 67 (1945), pp. 472-480. The present Note has been elaborated on the basis of a manuscript of Schur by M. Schiffer, Hebrew University, Jerusalem, in cooperation with Prof. M. Fekete.

pendence of $\psi(x)$ of the index n . Thus, Theorems I and II follow easily from Theorem III which will now be derived from Cauchy's Theorem.²

In fact, we have for small enough r

$$(7) \quad a_{n-\nu, \nu} = 1/2\pi i \int_{|\xi|=r} \frac{[g(\xi)]^{n-\nu}}{\xi^{\nu+1}} d\xi,$$

whence, in view of (3), for $|x|$ small enough

$$(8) \quad \begin{aligned} \phi_n(x) &= 1/2\pi i \int_{|\xi|=r} \frac{[g(\xi)]^n}{\xi} \cdot \sum_{\nu=0}^{\infty} \frac{x^\nu}{\xi^\nu [g(\xi)]^\nu} d\xi \\ &= 1/2\pi i \int_{|\xi|=r} \frac{g(\xi)^{n+1}}{\xi g(\xi) - x} d\xi. \end{aligned}$$

Consider now the equation

$$(9) \quad x = yg(y) = f(y)$$

which possesses for small enough $|x|$ a unique solution y in the neighborhood of $y=0$. Obviously, $\phi_n(x)$ is equal to the residue of the integrand of (8) at the point $\xi=y$. Hence, we get finally

$$(10) \quad \phi_n(x) = \phi_n[f(y)] = \frac{[g(y)]^{n+1}}{f'(y)},$$

which proves Theorem III.

Let us write the identity (5) in the form

$$(5') \quad x\psi[f(x)] = f(x).$$

This is equivalent to the theorem

IV. The function $y=f(x)=xg(x)$ has the inverse function

$$(11) \quad x=h(y)=y/\psi(y).$$

Another simple representation for the inverse function $h(y)$ is easily derived from (6). In fact, putting $n=-1$, we get

$$(6') \quad f'(x)[1 + a_{-2,1}f(x) + a_{-3,2}f(x)^2 + \cdots + a_{-k-1,k}f(x)^k + \cdots] = 1.$$

Integrating with respect to x between the limits 0 and x , we get identically in x

$$(12) \quad x = \sum_{\nu=1}^{\infty} (a_{-\nu, \nu-1}/\nu) f(x)^\nu.$$

V. The inverse function $x=h(y)$ of $y=f(x)$ has the development

² The following proof is due to Mr. George Schur.

$$(12') \quad x = h(y) = \sum_{\nu=1}^{\infty} (a_{-\nu, \nu-1}/\nu) y^{\nu}.$$

The calculation of the coefficients $a_{-\nu, \mu}$ is very much facilitated by the following remark. The coefficients $a_{-1, \mu}$ in the development

$$(13) \quad [g(x)]^{-1} = \sum_{\mu=0}^{\infty} a_{-1, \mu} x^{\mu}$$

are given by the well known formula

$$(13') \quad a_{-1, \mu} = \sum_{\alpha_1+2\alpha_2+\dots+\mu\alpha_{\mu}=\mu} (-1)^{\alpha_1+\alpha_2+\dots+\alpha_{\mu}} \frac{(\alpha_1 + \alpha_2 + \dots + \alpha_{\mu})!}{\alpha_1! \alpha_2! \dots \alpha_{\mu}!} a_1^{\alpha_1} a_2^{\alpha_2} \dots a_{\mu}^{\alpha_{\mu}}.$$

If, on the other hand, we differentiate identity (2) with respect to a_k and compare the coefficients of x^{μ} on both sides, we find

$$(14) \quad \partial a_{n\mu} / \partial a_1 = -n a_{n-1, \mu-k} \quad (\mu = k, k+1, \dots).$$

Thus, by means of (14), we may compute all coefficients $a_{-\nu, \mu}$ by differentiating the $a_{-1, \mu}$. In fact, we obtain

$$(15) \quad \partial a_{-n, \mu+1} / \partial a_1 = -n a_{-n-1, \mu}$$

whence, by iteration,

$$(15') \quad a_{-n, \mu} = \frac{(-1)^{n-1}}{(n-1)!} \frac{\partial^{n-1} a_{-1, \mu+n-1}}{\partial a_1^{n-1}}$$

The coefficients in the development (12') have, therefore, the form

$$(15'') \quad \frac{a_{-\nu, \nu-1}}{\nu} = \frac{(-1)^{\nu-1}}{\nu!} \frac{\partial^{\nu-1} a_{-1, 2(\nu-1)}}{\partial a_1^{\nu-1}}$$

and may be easily calculated from (13').

We further point out the following identity which results immediately from (11) and (12'):

$$(16) \quad \psi(x) \cdot \sum_{\nu=1}^{\infty} (a_{-\nu, \nu-1}/\nu) x^{\nu-1} = 1.$$

The identity (12') is closely related to Lagrange's inversion formula. This formula permits to solve the equation

$$(17) \quad x = a + yg(x)$$

for any given function $g(x)$, analytic at $x = a$, for small enough y . If $\phi(x)$ is a function which is analytic at $x = a$, we obtain, by the formula mentioned above, the following development of $\phi(x)$ into a series in powers of y :

$$(18) \quad \phi(x) = \phi(a) + \sum_{\nu=1}^{\infty} y^{\nu}/\nu! [(d^{\nu-1}/dx^{\nu-1})(\phi'(x)g(x)^{\nu})]_{x=a}$$

which converges for small enough y .

Let us put in (18) $a=0$, $\phi(x)=x$ and use our notations from (2). We obtain

$$(19) \quad x = \sum_{\nu=1}^{\infty} (a_{\nu, \nu-1}/\nu) y^{\nu}$$

as an inversion formula for

$$(19') \quad x = yg(x).$$

If we replace $g(x)$ by $g(x) = g(x)^{-1}$ we have to put instead of $a_{\nu, \nu-1}$ the coefficient $a_{-\nu, \nu-1}$. Thus, the inversion of $y = xg(x)$ is given exactly by (19'). This identity is, therefore, a simple consequence of Lagrange's formula.

Next, we apply Lagrange's formula (18) with $a=0$, $\phi(x)=x^k$ ($k=1, 2, \dots$). We find

$$(20) \quad x^k/k = \sum_{\nu=0}^{\infty} a_{k+\nu, \nu} (y^{k+\nu}/(k+\nu)).$$

Putting $x = yg(x)$, we immediately obtain

$$(20') \quad [g(x)]^k/k = \sum_{\nu=0}^{\infty} a_{-(k+\nu), \nu} (y^{\nu}/(k+\nu)).$$

Let us replace here again $g(x)$ by $g(x)^{-1}$ and at the same time $a_{\nu, \mu}$ by $a_{-\nu, \mu}$. We obtain the identity

$$(21) \quad [g(x)]^{-k}/k = \sum_{\nu=0}^{\infty} a_{-(k+\nu), \nu} (y^{\nu}/(k+\nu))$$

in x and y , if these variables are connected by the equation

$$(21') \quad y = xg(x) = f(x).$$

Multiplying both sides of (21) by $y^k = x^k g(x)^k$, leads to

$$(21'') \quad x^k/k = \sum_{\nu=0}^{\infty} a_{-(k+\nu), \nu} (y^{k+\nu}/(k+\nu)), \quad (k=1, 2, \dots).$$

Differentiating the last identity with respect to x , we obtain, in virtue of (3) and (21')

$$(22) \quad x^{k-1} = \sum_{\nu=0}^{\infty} a_{-(k+\nu), \nu} y^{k+\nu-1} \cdot f'(x) = y^{k-1} \cdot f'(x) \phi_{-k}[f(x)]$$

whence in view of (21')

$$(22') \quad [g(x)]^{-k+1} = f'(x) \phi_{-k}[f(x)].$$

Thus, we have derived identity (6) for n negative from Lagrange's formula.

2. Inverse Matrices. Lagrange's formula as well as our identity (6) may be proved by means of Cauchy's Theorem. The main task of this paper is to prove Theorem III in a purely arithmetic way and so open a new access to Lagrange's formula.

The decisive role in the establishment of the basic identities, used in this proof, is played by the following trivial

LEMMA. Denote by $[\psi(x)]_k$ the coefficient of x^k in the development of the arbitrary analytic function $\psi(x)$ into powers of x . With this notation, the function (1) satisfies for $k = 1, 2, \dots$ the equations

$$(23) \quad [g(x)^k + xg(x)^{k-1}g'(x)]_k = 0,$$

$$(23') \quad [g(x)^{-k} + xg(x)^{-k-1}g'(x)]_k = 0.$$

The truth of this lemma follows immediately from the fact that

$$(24) \quad g(x)^{k-1}g'(x) = (1/k)(g(x)^k)', \quad g(x)^{-k-1}g'(x) = (-1/k)(g(x)^{-k})'$$

whence

$$(24') \quad [xg(x)^{k-1}g'(x)]_k = a_{k,k}, \quad [xg(x)^{-k-1}g'(x)]_k = -a_{-k,k}$$

which, in virtue of (2), proves (23) and (23').

By means of the preceding lemma, we prove now the following

THEOREM. For every function $g(x)$ of the type (6) define the two matrices

$$(25) \quad M = (a_{\mu, \mu-\nu}) = \begin{bmatrix} 1 & 0 & 0 & \dots \\ a_{21} & 1 & 0 & \dots \\ a_{32} & a_{31} & 1 & \dots \\ \cdot & \cdot & \cdot & \dots \\ \cdot & \cdot & \cdot & \dots \end{bmatrix}, \quad N = (a_{\nu, \mu-\nu}) = \begin{bmatrix} 1 & 0 & 0 & \dots \\ a_{11} & 1 & 0 & \dots \\ a_{12} & a_{21} & 1 & \dots \\ \cdot & \cdot & \cdot & \dots \end{bmatrix}$$

with a_{np} vanishing for $p < 0$ and being defined by (2) for $p \geq 0$. Then, the inverse matrices are given by the formulae

$$(26) \quad M^{-1} = ((\mu/\nu)a_{-\nu, \mu-\nu}), \quad N^{-1} = ((\nu/\mu)a_{-\mu, \mu-\nu}).$$

Proof. Obviously, all four matrices mentioned in the theorem are triangular matrices, the diagonals of which consist of units. To prove, therefore, that M and M^{-1} are in fact inverse matrices, it suffices to show that for $\mu > \nu \geq 1$ we have

$$(27) \quad J_{\mu\nu} = \sum_{\nu \leq \lambda \leq \mu} a_{\mu, \mu-\lambda} \cdot (\lambda/\nu)a_{-\nu, \lambda-\nu} = 0.$$

Now, it is easily verified that

$$(28) \quad [g(x)^\mu \cdot (x^\nu g(x)^{-\nu})' / x^{\nu-1}]_{\mu-\nu} = \sum_{\nu \leq \lambda \leq \mu} a_{\mu, \mu-\lambda} \lambda a_{-\nu, \lambda-\nu} = \nu J_{\mu\nu}$$

and that, therefore,

$$(28') \quad \begin{aligned} \nu J_{\mu\nu} &= [\nu g(x)^\mu (g(x)^{-\nu} - xg(x)^{-\nu-1}g'(x))]_{\mu-\nu} \\ &= \nu [g(x)^{\mu-\nu} - xg(x)^{\mu-\nu-1}g'(x)]_{\mu-\nu} = 0 \end{aligned}$$

because of (23). Thus, we have proved (27).

In exactly the same way, we may prove for every $\mu > \nu \geq 1$ that

$$(29) \quad J_{\mu\nu} = \sum_{\nu \leq \lambda \leq \mu} (\lambda/\mu) a_{-\mu, \mu-\lambda} \cdot a''_{\lambda, \lambda-\nu} = 0$$

which is equivalent to the matrix equation $N^{-1} \cdot N = 1$. In fact, we have

$$(30) \quad [g(x)^{-\mu} ((x^\nu g(x)^\nu)' / x^{\nu-1})]_{\mu-\nu} = \sum_{\nu \leq \lambda \leq \mu} \lambda a_{-\mu, \mu-\lambda} a_{\nu, \lambda-\nu} = \mu J_{\mu\nu}.$$

On the other hand, we may write the left hand expression in the form

$$(30') \quad \begin{aligned} [\nu g(x)^{-\mu} (g(x)^\nu + xg(x)^{\nu-1}g'(x))]_{\mu-\nu} \\ = \nu [g(x)^{-(\mu-\nu)} + xg(x)^{-(\mu-\nu)-1}g'(x)]_{\mu-\nu}, \end{aligned}$$

which vanishes in view of (23'). Hence, $y_{\mu\nu} = 0$, which completes our proof.

We introduce a further matrix

$$(31) \quad K = (\mu \delta_{\mu\nu}) = \begin{bmatrix} 1 & 0 & 0 & 0 & \cdot \\ 0 & 2 & 0 & 0 & \cdot \\ 0 & 0 & 3 & 0 & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \end{bmatrix}$$

and write $M = M(g)$, $N = N(g)$ in order to emphasize the dependence of these matrices on the function $g(x)$. Then, we obviously have in view of (25) and (26) the matrix equation

$$(32) \quad M^{-1}(g) = KN(1/g)K^{-1},$$

since replacing g by $1/g$ means a change from $a_{\nu, \mu}$ to $a_{-\nu, \mu}$ in all matrices. Hence,

$$(33) \quad K = M(g) \cdot K \cdot N(1/g)$$

or replacing g by $1/g$,

$$(33') \quad K = M(1/g) \cdot K \cdot N(g)$$

i. e.,

$$(33'') \quad N(g)^{-1} = K^{-1}M(1/g)K = ((\nu/\mu)a_{-\mu, \mu-\nu}).$$

Thus, we see that the knowledge of M^{-1} , leading to the identity (32), provides a new way for calculating N^{-1} .

3. Lagrange's Inversion Formula. The matrix identities found in the last paragraph will be applied now to prove Lagrange's formula—and thus all its consequences—in a purely algebraic way.

First of all, we want to solve the equation

$$(21') \quad y = xg(x)$$

by means of a series in powers of y

$$(34) \quad x = y + A_1 y^2 + A_2 y^3 + \cdots$$

In order to determine all A , insert in (34) $y = xg(x)$ and compare on both sides the coefficient of x^m . Putting $\delta_{im} = \begin{cases} 1 & \text{for } m = i \\ 0 & \text{for } m \neq i \end{cases}$ we get

$$(35) \quad \delta_{im} = a_{1,m-1} + A_1 a_{2,m-2} + A_2 a_{3,m-3} + \cdots + A_{m-1} a_{m,0}, \quad (m = 1, 2, \cdots).$$

$$(35') \quad \delta = \begin{bmatrix} \delta_{11} \\ \delta_{12} \\ \cdot \\ \cdot \\ \cdot \\ \cdot \end{bmatrix}, \quad A = \begin{bmatrix} 1 \\ A_1 \\ A_2 \\ \cdot \\ \cdot \\ \cdot \end{bmatrix};$$

in this notation (35) may be written as

$$(35'') \quad \delta = N \cdot A$$

which yields by inversion

$$(36) \quad A = N^{-1} \cdot \delta,$$

whence, in view of (26) and (35'),

$$(37) \quad A_\mu = (1/\mu + 1) a_{-(\mu+1), \mu}.$$

Thus, we have solved the equation (21') uniquely by the development

$$(34') \quad x = y + a_{-2,1}(y^2/2) + a_{-3,2}(y^3/3) + \cdots$$

which is exactly the statement of Theorem V.

Next, let us deal with the more general question of determining the coefficients in the development

$$(38) \quad x^k = B_k y^k + B_{k+1} y^{k+1} + \dots$$

where y and x are connected by the equation (21'). Comparing on both sides the coefficient of x^m , we obtain

$$(39) \quad \delta_{km} = B_k A_{k, m-k} + B_{k+1} A_{k+1, m-k-1} + \dots + B_m a_{m, 0}.$$

In order to express (39) in the form of a vector equation, we define the vectors

$$(39') \quad \delta_k = \begin{bmatrix} \delta_{k1} \\ \delta_{k2} \\ \vdots \\ \delta_k \\ \vdots \end{bmatrix}, \quad B = \begin{bmatrix} B_1 \\ B_2 \\ \vdots \\ B_k \\ \vdots \end{bmatrix} \quad \text{with} \quad B_1 = \dots B_{k-1} = 0, \quad B_k = 1, \dots$$

By means of these definitions we may write instead of (39)

$$(39'') \quad \delta_k = N \cdot B$$

whence, on multiplication by N^{-1} ,

$$(40) \quad B = N^{-1} \cdot \delta_k \quad \text{i. e.,} \quad B_\mu = (k/\mu) a_{-\mu, \mu-k}.$$

The development (38) has, therefore, the form

$$(41) \quad x^k = y^k + (k/(k+1)) a_{-(k+1), 1} y^{k+1} \\ + (k/(k+2)) a_{-(k+2), 2} y^{k+2} + \dots, \quad k = 2, 3, \dots$$

The formula (41), valid in view of (34') also for $k=1$, is identical with (21''), i. e., with Lagrange's inversion formula for the equation $y = xg(x)$. We have derived here this important formula in a purely arithmetic way.

It is usual to give the inversion formula for the equation

$$(42) \quad x = yg(x).$$

We have already dealt in 1 with the connections of the corresponding results (compare formula (20)).

We have derived already in 1 the identity (6), for negative n , from equation (21) in a purely formal way. Hence, having proved (21) by arithmetic means only, we did the same for identity (6) in case of n negative. But this identity may be conceived as an identity between the coefficients of x^n on both sides, which are polynomials in n, a_1, a_2, \dots . Therefore, these identities remain valid for n positive also.

114708

We may even generalise this last result. We put

$$(43) \quad \gamma(x) = a_1x + a_2x^2 + \dots, \quad g(x) = 1 + \gamma(x)$$

and define for *every* real or complex κ

$$(44) \quad g(x)^\kappa = \sum_{\nu=0}^{\infty} \binom{\kappa}{\nu} \gamma(x)^\nu = \sum_{\nu=0}^{\infty} a_{\kappa, \nu} x^\nu, \quad \phi_\kappa(x) = \sum_{\nu=0}^{\infty} a_{\kappa, \nu} x^\nu.$$

Evidently, the $a_{\kappa, \nu}$ are still polynomials in κ, a_1, a_2, \dots . Hence, the identity (6), being valid for $n = -1, -2, -3, \dots$, remains valid for all complex values of n . Thus, we have proved the most general form of (6) by purely algebraic means.

Finally, we mention another group of identities. In analogy to $\phi_n(x)$ we define series

$$(45) \quad \psi_n(x) = \sum_{\nu=0}^{\infty} a_{n+\nu, \nu} x^\nu.$$

We emphasize the correspondence between ϕ_n, ψ_n and the initial function $g(x)$ by the notation $\phi_n(x; g)$ and $\psi_n(x; g)$. Then, we have obviously

$$(46) \quad \psi_n(x; g) = \phi_{-n}(x; (1/g)).$$

For every integer value of n we have, in view of (16),

$$(47) \quad g(x)^{n+1} = (xg(x))' \phi_n(xg(x); g).$$

Replacing $g(x)$ by $1/g(x)$ yields, in virtue of (46),

$$(47') \quad g(x)^{-(n+1)} = (x/g(x))' \phi_n(x/g(x); 1/g) = (x/g(x))' \psi_{-n}(x/g(x); g).$$

Writing $-n$ instead of n yields clearly

$$(47'') \quad g(x)^{n+1} = E(x) \psi_n(x/g(x); g), \quad E(x) = g(x) - xg'(x).$$

From (47'') it is obvious that

$$(48) \quad \psi(x) = \psi_{n+1}(x)/\psi_n(x)$$

is independent of n ; but this result is also a consequence of definition (46) and Theorem I.

Further, we may write, in view of (47') and (47''),

$$(49) \quad g(x)^{n+1} = (g(x) + xg'(x)) \psi_n(xg(x)) = (g(x) - xg'(x)) \psi_n(x/g(x)),$$

whence the result that

$$(49') \quad \psi_n(x/g(x))/\psi^n(xg(x)) = g(x) + xg'(x)/g(x) - xg'(x)$$

is independent of n .

4. Another Inversion Formula. In this paragraph we shall consider series of the type

$$(50) \quad w = z + a_1 + a_2/z + a_3/z^2 + \dots = zg(1/z),$$

$$g(x) = 1 + a_1x + a_2x^2 + \dots$$

We wish to invert (50) into the form

$$(51) \quad z = w + A_1 + A_2/w + \dots$$

In order to determine the coefficients A_v , put $x = 1/z$ and consider the identity in z :

$$(51') \quad z = zg(x) + A_1 + A_2x/g(x) + A_3x^2/g(x)^2 + \dots$$

Comparing the coefficients of x^v on both sides of this identity yields

$$(52) \quad A_1 = -a_1, \quad a_v + \sum_{\rho=1}^{v-1} A_{\rho+1}a_{-\rho, v-\rho-1} = 0 \quad (v = 2, 3, \dots).$$

We introduce the vectors

$$(52') \quad C = \begin{bmatrix} -a_2 \\ -a_3 \\ -a_4 \\ \vdots \\ \vdots \\ \vdots \end{bmatrix}, \quad D = \begin{bmatrix} A_2 \\ A_3 \\ \vdots \\ \vdots \\ \vdots \end{bmatrix};$$

then, the system of equations of (52) may be written, in view of definition (25), in a single vector formula

$$(52'') \quad C = N(1/g) \cdot D,$$

whence, in virtue of (26),

$$(53) \quad D = N(1/g)^{-1} \cdot C, \quad \text{i. e., } \mu A_{\mu+1} = - \sum_{\lambda=1}^{\mu} \lambda a_{\mu, \mu-\lambda} \cdot a_{\lambda+1}.$$

We introduce the series

$$(54) \quad \phi(x) = a_2x^2 + 2a_3x^3 + \dots + na_{n+1}x^{n+1} + \dots$$

We have by (53) evidently

$$(55) \quad [g(x)^n \cdot \phi(x)]_{n+1} = \sum_{\lambda=1}^n \lambda a_{n, n-\lambda} \cdot a_{\lambda+1} = -nA_{n+1}.$$

On the other hand,

$$(56) \quad g(x) + \phi(x) - 1 = a_1x + 2a_2x^2 + 3a_3x^3 + \dots = xg'(x);$$

hence, in virtue of (55) and (56),

$$(57) \quad -nA_{n+1} = [xg(x)^n \cdot g'(x) - g(x)^{n+1} + g(x)^n]_{n+1}.$$

Now, we apply the lemma of 2; in virtue of (23) with $k = n + 1$ we have

$$(57') \quad -nA_{n+1} = [g(x)^n]_{n+1} = a_{n,n+1}.$$

We find, therefore, for (51) the simple expression

$$(58) \quad z = w - a_1 - a_{1,2}/1 \cdot (1/w) - a_{2,3}/2 \cdot (1/w^2) - \dots - a_{n,n+1}/n \cdot (1/w^n) -$$

If we write $x = 1/z$, $y = 1/w$, we may put (50) into the form

$$(42') \quad x = yg(x).$$

This allows the following interpretation of (58):

$$(59) \quad -1/x = -1/y + a_1 + (a_{1,2}/1)y + (a_{2,3}/2)y^2 \\ + \dots + (a_{n,n+1}/n)y^n + \dots,$$

i. e., this series is a generalization of Lagrange's inversion formula (20) for $k = -1$. One might be tempted to apply directly (20) with $k = -1$; but in this formula appears the term $a_{k+1,1}/(k+1)$ which is undefined for $k = -1$. If, however, we replace this meaningless term by a_1 the formula (20) with $k = -1$ is the same, term by term, as (59).

Let us write (20) with $k = 1$; we obtain

$$(60) \quad x = y + (a_{2,1}/2)y^2 + (a_{3,2}/3)y^3 + \dots + (a_{n,n-1}/n)y^n + \dots$$

Multiplying (59) by (60) we obtain the identity

$$(61) \quad (1 - a_1y - (a_{1,2}/1)y^2 - (a_{2,3}/2)y^3 - \dots) \\ (1 + (a_{2,1}/2)y + (a_{3,2}/3)y^2 + \dots) = 1.$$

There arises now the question of generalizing Lagrange's formula (20) for general k . We proceed in an analogous way as in 3. We solve equation (42) by the development (60):

$$(62) \quad x = yh(y) = y(1 + (a_{2,1}/2)y + (a_{3,2}/3)y^2 + \dots).$$

We introduce this development into the formula (20), valid for positive integer values of k . After dividing by y^k we obtain

$$(63) \quad h(y)^k/k = \sum_{\nu=0}^{\infty} (a_{k+\nu,\nu}/k + \nu)y^\nu.$$

If we write

$$(64) \quad h(y)^n = \sum_{v=0}^{\infty} b_{n,v} y^v \quad (n = 0, \pm 1, \pm 2, \dots)$$

we get, by comparing the coefficients of like powers of y^m on both sides of (63),

$$(65) \quad b_{k,m}/k = a_{k+m,m}/k + m \quad (k = 1, 2, 3, \dots; m = 0, 1, 2, \dots).$$

By definitions (2), (62) and (64), the $a_{k+m,m}$ and $b_{k,m}$ are, for m fixed, polynomial functions of k, a_1, a_2, \dots . Therefore, on each side of (65) stands a rational function of k and both functions coincide for positive integer values of k . Hence, these functions are identical and the equality (65) holds for *every* (complex) value of k . Formula (20) is, therefore, also true for every k , except for negative integers for which one of its terms becomes meaningless.

In order to generalize (20) for the case of negative integers too, we have only to replace the undefined quotient $a_{k+n,n}/(k+n)$ in the case $k = -n$ by the limit

$$\lim_{k \rightarrow -n} a_{k+n,n}/(k+n).$$

We have obviously

$$(66) \quad \lim_{l=0} \frac{g(x)^l - 1}{l} = \log g(x),$$

whence, by comparing the coefficients of x^m on both sides,

$$(67) \quad \lim_{l=0} a_{l,m}/l = [\log g(x)]_m. \quad (m = 1, 2, \dots).$$

Hence, in the case $k = -n$, we have to replace in (20) the term $a_{k+n,n}/(k+n)$ by $[\log g(n)]_n$.

If we put

$$(68) \quad \log g(x) = \sum_{v=1}^{\infty} \gamma_v x^v$$

we may write Lagrange's inversion formula (20) for $k = -n$ in the form

$$(69) \quad \frac{x^{-n}}{-n} = \frac{y^{-n}}{-n} + (a_{-(n-1),1}/-(n-1))y^{-(n-1)} \\ + \dots + (a_{-1,n-1}/-1)y^{-1} + \gamma_n + (a_{1,n+1}/1)y + \dots$$

In a similar way, we obtain a new identity by adding to both sides of (63) the term $-1/k$ and then passing to the limit $k = 0$. We obtain

$$(70) \quad \log h(y) = \sum_{v=1}^{\infty} (a_{v,v}/v) y^v.$$

On the other hand, we have in view of $x = yg(x)$ and $x = yh(y)$ obviously $h(y) = g(x)$ and thus

$$(70') \quad \log g(x) = \sum_{\nu=1}^{\infty} (a_{\nu,\nu}/\nu) y^{\nu}.$$

Finally, we apply the formula (69) in the case of series (50). We put, as before, $x = 1/z$, $y = 1/w$ which transforms equation (50) into (42) and vice versa. Instead of (69), we obtain, for positive integers n ,

$$(71) \quad z^n/n = w^n/n + (a_{-(n-1),1}/(n-1))w^{(n-1)} + \cdots + (a_{-1,n-1}/1)w - \gamma_n \\ - a_{1,n+1}/1 \cdot (1/w) - a_{2,n+2}/2 \cdot (1/w^2) - \cdots$$

Consider now the polynomial

$$(72) \quad P_n(w) = n[w^n/n + (a_{-(n-1),1}/(n-1))w^{n-1} \\ + \cdots + (a_{-1,n-1}/1)w - \gamma_n];$$

it has the remarkable property that by introducing $w = zg(1/z) = f(z)$ we get a development

$$(72) \quad P_n(f(z)) = z^n + \alpha_1/z + \alpha_2/z^2 + \cdots,$$

with z^n as the only non-negative power of z . By this characteristic property the polynomial $P_n(w)$ is associated uniquely with the series $w = f(z)$ and is called the n -th Faber polynomial of $f(z)$. It plays an important role in the theory of conformal representation.³

On comparing the developments (68) and (70') we obtain, in view of (20), the equality

$$(73) \quad \sum_{k=1}^{\infty} \gamma_k x^k = \sum_{k=1}^{\infty} k \gamma_k \sum_{\nu=0}^{\infty} (a_{k+\nu,\nu}) y^{k+\nu}/k + \nu = \sum_{\nu=1}^{\infty} (a_{\nu,\nu}/\nu) y^{\nu}.$$

Comparing the coefficients of equal powers of y on both sides yields

$$(74) \quad \sum_{\nu=0}^{m-1} (m-\nu) \gamma_{m-\nu} a_{m,\nu} = a_{m,m} \quad (m = 1, 2, \cdots),$$

an interesting relation between the $a_{\mu\nu}$.

³ Compare: Schur, I., "On Faber Polynomials," *American Journal of Mathematics*, vol. 67 (1945), pp. 33-41.

A NOTE ON DIVISION RINGS.*

By N. JACOBSON.

In this note we consider division subrings Φ of a division ring P such that the left dimensionality $(P : \Phi)_l$ is finite. If \mathfrak{L} denotes the ring of linear transformations in P regarded as a vector space over the set Φ_l of left multiplications by elements of Φ then clearly \mathfrak{L} contains the subring P_r of right multiplications by elements of P . The relation $(\mathfrak{L} : P_r)_r = (P : \Phi)_l$ can be established (Theorem 1). Thus \mathfrak{L} has finite right dimensionality over P_r . The principal result of this paper is that the converse of these results holds, namely, if \mathfrak{A} is any ring of endomorphisms of the additive group of P such that 1) \mathfrak{A} contains P_r and 2) $(\mathfrak{A} : P_r)_r$ is finite, then \mathfrak{A} coincides with a ring \mathfrak{L} of linear transformations of P over Φ_l where Φ is a division subring of P such that P over Φ_l is finite. A complete reciprocity is established between the rings \mathfrak{A} (or \mathfrak{L}) and the subrings Φ of P .

There are a number of applications of our results. One of these, the theory of automorphisms in a division ring, is developed here. In this connection we generalize the concept of a closed group of automorphisms introduced by Emmy Noether¹ for groups of inner automorphisms and we establish a Galois correspondence between closed groups that satisfy a certain finiteness condition and their division subrings of invariants. The theorem of Noether²-Artin and Whaples³ on commutators of division subalgebras containing the center Γ and our own results on finite groups of outer automorphisms are contained as special cases of the theorems derived here.

1. Let P be a division ring and Φ a division subring of P of *finite left index* m . By this we mean that the left dimensionality $(P : \Phi)_l = m < \infty$. Here we are regarding P as a left vector space (or module) over Φ or, what amounts to the same thing, as a group relative to the set Φ_l of left multiplications $\xi \rightarrow \xi\alpha \equiv \alpha\xi$, ξ in P , α in Φ . Let \mathfrak{L} denote the ring of linear transformations (l. t.) of P over Φ_l . Thus $A \in \mathfrak{L}$ if and only if A is an endomorphism of the additive group of P and

* Received June 28, 1946.

¹ [5], p. 529.

² [5], p. 528.

³ [1], p. 104.

$$(\alpha\xi)A = \alpha(\xi A)$$

for all $\alpha \in \Phi$ and all $\xi \in P$. Evidently \mathfrak{Q} contains the set P_r of right multiplications $\xi \rightarrow \xi\rho \equiv \xi\rho$ in P . Since $1 \in P_r$, \mathfrak{Q} may be regarded as a (right) vector space over P_r . When this is done we obtain the following theorem:

THEOREM 1. *The dimensionalities $(\mathfrak{Q} : P_r)_r$ and $(P : \Phi)_l$ are equal.*⁴

Let $\xi_1, \xi_2, \dots, \xi_m$ be a left basis of P over Φ . Then any l. t. in P over Φ_l is completely determined by its effect on the ξ_i . Now define C_j , $j = 1, 2, \dots, m$, to be the l. t. such that

$$\xi_j C_j = 1 \quad \text{and} \quad \xi_i C_j = 0 \quad \text{if } i \neq j.$$

Then $\Sigma C_j \eta_{jr}$ maps ξ_i into η_i . Hence $\Sigma C_j \eta_{jr} = 0$ implies that all the $\eta_j = 0$. Thus the C 's are right independent relative to P_r . On the other hand if A is any l. t. in P over Φ_l , let $\xi_i A = \eta_i$. Then A coincides with $\Sigma C_j \eta_{jr}$. Hence the C 's form a right basis for \mathfrak{Q} over P_r .

2. We assume now that \mathfrak{A} is any ring of endomorphisms acting in the additive group of P that 1) contains P_r and 2) has finite dimensionality $(\mathfrak{A} : P_r)_r$. Let C be any endomorphism in P that commutes with every $A \in \mathfrak{A}$. Then since C commutes with every right multiplication, $C = \gamma_l$ a left multiplication.⁵ The totality of γ obtained in this way is a division subring Φ of P that we shall call the division ring of \mathfrak{A} -invariants.

THEOREM 2. *Let \mathfrak{A} be a ring of endomorphisms of the additive group of P such that 1) $\mathfrak{A} \supseteq P_r$ and 2) $(\mathfrak{A} : P_r)_r = m < \infty$ and let Φ be the division subring of \mathfrak{A} -invariants. Then $(P : \Phi)_l = m$ and $\mathfrak{A} = \mathfrak{Q}$ the complete ring of l. t. of P over Φ_l .*

Since $\mathfrak{A} \supseteq P_r$, \mathfrak{A} is an irreducible ring of endomorphisms in P . Since $(\mathfrak{A} : P_r)_r < \infty$, \mathfrak{A} satisfies the descending chain condition for right ideals. Now Φ_l is the division ring of endomorphisms commutative with the elements of the irreducible ring \mathfrak{A} . It follows from known results on irreducible rings of endomorphisms that $(P : \Phi)_l < \infty$ and that $\mathfrak{A} = \mathfrak{Q}$ the complete ring of l. t. of P over Φ_l .⁶ By Theorem 1 $(P : \Phi)_l = (\mathfrak{A} : P_r)_r$.

To complete the reciprocity between Φ and \mathfrak{Q} we require the following theorem:

⁴ The present proof of this theorem, which is simpler than my original one, was communicated to me by R. Brauer.

⁵ [2], p. 54.

⁶ [4], in particular Theorems 6 and 3.

THEOREM 3. *Let Φ be a division subring of finite left index P and let \mathfrak{Q} be the ring of l. t. in P over Φ_l . Then reciprocally Φ_l is the complete set of endomorphisms in P that commute with all the endomorphisms belonging to \mathfrak{Q} .*

This result is well known for arbitrary vector spaces.⁷ It can also be obtained as a consequence of Theorems 1 and 2. For let Ψ be the division ring of \mathfrak{Q} -invariants. Then by Theorem 2, $(P : \Psi)_l = (\mathfrak{Q} : P_r)_r$. On the other hand $\Psi \supseteq \Phi$ and $(P : \Phi)_l = (\mathfrak{Q} : P_r)_r$ by Theorem 1. Hence $\Psi = \Phi$.

If $(P : \Phi)_l < \infty$ and \mathfrak{Q} is the ring of l. t. of P over Φ_l our results establish a $(1-1)$ correspondence between the division rings Ψ between P and Φ and the rings \mathfrak{B} between P_r and \mathfrak{Q} . If Ψ is given then \mathfrak{B} is the ring of l. t. of P over Ψ_l . Clearly $P_r \subseteq \mathfrak{B} \subseteq \mathfrak{Q}$. On the other hand if \mathfrak{B} is a ring between P_r and \mathfrak{Q} then the set Ψ of \mathfrak{B} -invariants is a division subring containing Φ . These two correspondences are inverses of each other.

All of the above results can, of course, be duplicated for division subrings of finite right index. In the remainder of this note we consider some applications of these theorems to the theory of automorphisms in a division ring.

3. Let A be an automorphism in P . Then A is an endomorphism of the additive group of P and $(\xi\eta)A = (\xi A)(\eta A)$. Thus we have the following fundamental commutation rules governing automorphisms and multiplications:

$$(2) \quad \eta_r A = A(\eta A)_r \quad \xi_l A = A(\xi A)_l.$$

We consider now the possible forms of linear relations with coefficients that are multiplications that can hold for automorphisms in P . First let I_1, I_2, \dots, I_h be inner automorphisms, say, $\xi I_j = \tau_j \xi \tau_j^{-1}$. Then $I_j = \tau_{j,l} \tau_{j,r}^{-1}$ if $\tau_{j,l}$ is the left multiplication determined by τ_j and $\tau_{j,r}$ is the right multiplication determined by τ_j . Suppose that the τ 's are linearly dependent over Γ where Γ is the center of P . Then there exist γ_i not all 0 such that $\sum \tau_i \gamma_i = 0$. Hence $\sum I_i \tau_{i,r} \gamma_{i,r} = \sum \tau_{i,l} \gamma_{i,l} = \sum \tau_{i,l} \gamma_{i,l} = 0$. Since not every $\tau_{j,r} \gamma_{j,r}$ equals 0 this shows that the I_j are linearly dependent over P_r .

Conversely suppose that we have a relation $\sum I_i \eta_{i,r} = 0$ where not every $\eta_{j,r} = 0$. Then $\sum \tau_{i,l} \mu_{i,r} = 0$ where $\mu_j = \tau_j^{-1} \eta_j$ so that not every $\mu_j = 0$. We assert that this implies that the τ_j are dependent over Γ . This is the following well-known lemma:

LEMMA 1. *If $\tau_1, \tau_2, \dots, \tau_h$ are elements of P that are linearly independent over Γ then the left multiplications $\tau_{1,l}, \tau_{2,l}, \dots, \tau_{h,l}$ are linearly independent over P_r .*

⁷ [2], p. 22.

For the sake of completeness we give the proof. If the τ_i 's are dependent over P_r there exists a shortest relation which we can suppose to be of the form

$$(3) \quad \tau_{1l}\mu_{1r} + \tau_{2l}\mu_{2r} + \cdots + \tau_{sl}\mu_{sr} = 0$$

where all the μ_i are different from 0. We may assume also that $\mu_{1r} = 1_r$. Since the τ_j are independent over Γ , one of the μ 's, say μ_2 , is not in Γ . Hence there exists a ν such that $\mu_2\nu - \nu\mu_2 \neq 0$. Multiplication of (3) on the left and on the right by ν_r followed by subtraction of the resulting relations gives a new relation shorter than (3). This contradiction proves the result.

We suppose next that A_1, A_2, \dots, A_u are automorphisms that are in distinct cosets of the subgroup of inner automorphisms. Then no $A_i A_j^{-1}$, $i \neq j$, is inner. Let \mathcal{U} denote the ring of endomorphisms generated by the left and the right multiplications. The elements of \mathcal{U} have the form $U = \tau_{1l}\mu_{1r} + \tau_{2l}\mu_{2r} + \cdots$. We wish to prove the following lemma.

LEMMA 2. *If A_1, A_2, \dots, A_u are automorphisms such that no $A_i A_j^{-1}$, $i \neq j$, is inner, then a relation $\sum A_i U_i = 0$ where the $U_j \in \mathcal{U}$ can hold only if every $U_j = 0$.*

We may choose elements $\tau_1, \tau_2, \dots, \tau_h$ such that the left multiplications τ_{il} are independent over P_r and such that every $U_i = \sum \tau_{jl}\mu_{jlr}$. Then our relation reads $\sum A_i \tau_{jl}\mu_{jlr} = 0$. We shall show that every $\mu_{ji} = 0$. For otherwise we may suppose that $\sum A_i \tau_{jl}\mu_{jlr} = 0$ is a shortest linear relation connecting the hu elements $A_i \tau_{jl}$ in which the coefficients $\mu_{jlr} \neq 0$. We may suppose also that $\mu_{11} = 1$. We shall show that no A other than A_1 occurs in this relation. For suppose that

$$(4) \quad A_1 \tau_{1l}\mu_{11r} + A_2 \tau_{kl}\mu_{k2r} + \cdots = 0.$$

Multiply (4) on the left by ξ_r and on the right by $(\xi A_1)_r$ and subtract the resulting relations. This yields

$$(5) \quad A_2 \tau_{kl} [(\xi A_2)_r \mu_{k2r} - \mu_{k2r} (\xi A_1)_r] + \cdots = 0.$$

If the bracket is 0 for all ξ

$$\xi A_2 = \mu_{k2} (\xi A_1) \mu_{k2}^{-1}$$

contrary to the assumption that A_1 and A_2 do not differ by an inner automorphism. Hence there is a ξ such that the bracket is not 0 and this gives a shorter relation than (4). Thus (4) must have the form

$$A_1 (\tau_{1l}\mu_{11r} + \cdots) = 0.$$

Since A_1 is an automorphism this implies that $\tau_{11}\mu_{11r} + \dots = 0$ contrary to the assumption that the τ_{ji} are linearly independent over P_r .

Suppose finally that we have an arbitrary linear relation with coefficients that are in P_r connecting a finite set of automorphisms. We may write this in the form $\sum A_i I_j \mu_{jir} = 0$ where the I_j are inner and the A_i are in different cosets relative to the subgroup of inner automorphisms. If $I_j = \tau_{ji}\tau_{jr}^{-1}$, $\sum A_i \tau_{ji}\tau_{jr}^{-1} \mu_{jir} = 0$. Hence by Lemma 2, $\sum I_j \mu_{jir} = \sum \tau_{ji}\tau_{jr}^{-1} \mu_{jir} = 0$. By Lemma 1, either all the μ 's are 0 or the τ_j are linearly dependent over Γ .

4. An element ϕ of P is *invariant* under an automorphism A if $\phi A = \phi$. By (2) ϕ has this property if and only if the left (right) multiplication $\phi_l(\phi_r)$ commutes with A . If G is a group of automorphisms we shall call ϕ a *G-invariant* if ϕ is invariant under every $A \in G$. The totality of these elements is a division subring Φ of P . Let $\mathfrak{A} = GP_r$ denote the totality of endomorphisms in the additive group of P that have the form $\sum A_i \eta_{ir}$ where the $A_i \in G$. Clearly \mathfrak{A} is closed under addition and subtraction and by the first equation of (2) \mathfrak{A} is closed under multiplication. Hence \mathfrak{A} is a subring of the ring of endomorphisms of P . Since G contains the identity, \mathfrak{A} contains P_r . Thus an endomorphism commutes with every element of \mathfrak{A} if and only if it is a left multiplication that commutes with every $A \in G$. It follows that the division ring of G -invariants coincides with the set of \mathfrak{A} -invariants in the sense defined above.

Let H be the invariant subgroup of inner automorphisms contained in G and let $I_1 = \tau_{11}\tau_{1r}^{-1}$ and $I_2 = \tau_{21}\tau_{2r}^{-1}$ belong to H . Then clearly $I_3 = I_1 I_2 = \tau_{31}\tau_{3r}^{-1}$ where $\tau_3 = \tau_1 \tau_2$. An element ϕ is an I_1 and an I_2 -invariant if and only if it commutes with τ_1 and with τ_2 . If this is so, ϕ also commutes with every $\tau = \tau_1 \gamma_1 + \tau_2 \gamma_2$, γ_i in Γ . Hence we do not change the set of G -invariants if we adjoin to G the inner automorphisms by the element τ obtained in this way. This remark leads us to restrict our attention to the study of groups G that are *closed* in the sense that if $I_1 = \tau_{11}\tau_{1r}^{-1}$ and $I_2 = \tau_{21}\tau_{2r}^{-1}$ are in G then so is $I = \tau_{11}\tau_{1r}^{-1}$, where $\tau = \tau_1 \gamma_1 + \tau_2 \gamma_2$ and the γ_i are arbitrary in Γ . It is readily seen that G is closed if and only if the subgroup H is precisely the set of inner automorphisms that are defined by the non-zero elements of the division subalgebra A of P that contains the center Γ . The algebra A , which is uniquely determined, will be referred to as the division subalgebra *associated* with the closed group G (or H).

We shall say that a closed group G has finite reduced order if 1) $(A : \Gamma)$ is finite and 2) H has finite index in G . The product of these two numbers will be called the *reduced order* of G . Our aim in the remainder

of this note is the establishment of a Galois correspondence between closed groups of automorphisms of finite reduced order and their division rings of invariants. We prove first the following theorem:

THEOREM 4. *Let G be a closed group of automorphisms of finite reduced order in the division ring P and let $\mathfrak{A} = GP_r$ be the ring of endomorphisms of the form $\sum A_i \eta_i r$, A_i in G . Then $(\mathfrak{A} : P_r)_r = \text{reduced order of } G$.*

Let A_1, \dots, A_h be representatives of the cosets of H in G and let I_1, \dots, I_h be inner automorphisms such that if $I_j = \tau_j I \tau_j^{-1}$ then the elements τ_j form a basis for the subalgebra A associated with G . Then any automorphism in G has the form $A_k I$ where I is in H . If $I = \tau_i I \tau_i^{-1}$, $\tau = \sum \tau_i \gamma_i$. As we have seen this implies that $I = \sum I_j \eta_j r$ where $\eta_j = \tau_j \gamma_j \tau^{-1}$. Hence $A_k I$ is a linear combination of the automorphisms $A_k I_1, A_k I_2, \dots, A_k I_h$. Consequently the $h\mu$ automorphisms $A_i I_j$ are generators over P_r of \mathfrak{A} . Now if $\sum A_i I_j \mu_{jir} = 0$ then $\sum I_j \mu_{jir} = 0$ by Lemma 2. Moreover $\sum I_j \mu_{jir} = 0$ and the independence of the τ_j over Γ implies that all the $\mu_{ji} = 0$. Thus the $A_i I_j$ are independent over P_r . Hence $(\mathfrak{A} : P_r)_r = h\mu$.

5. We may now establish the Galois correspondence between groups of automorphisms and their division rings of invariants. The basic results are contained in Theorems 5 and 8.

THEOREM 5. *Let G be a closed group of automorphisms of finite reduced order in the division ring P and let Φ be the division subring of G -invariants. Then $(P : \Phi)_l = \text{reduced order of } G = (P : \Phi)_r$ and any automorphism of P that leaves the elements of Φ invariant belongs to G .*

Let $\mathfrak{A} = GP_r$. Then by Theorem 4, $(\mathfrak{A} : P_r)_r = \text{reduced order of } G$. On the other hand we know that Φ is the division ring of \mathfrak{A} -invariants. Hence by Theorem 2, $(P : \Phi)_l = (\mathfrak{A} : P_r)_r$. Hence $(P : \Phi)_l = \text{reduced order of } G$. The equality $(P : \Phi)_r = \text{reduced order of } G$ follows by symmetry since the concept of reduced order is a "two-sided" one. Suppose now that A is any automorphism in P with the property that $\phi A = \phi$ for all $\phi \in \Phi$. Then $A\phi_l = \phi_l A$ and A is a linear transformation in P over Φ_l . Hence by Theorem 2, $A \in \mathfrak{A}$. Thus $A = \sum A_i I_j \mu_{jir}$ where the A_i and I_j are determined as in the proof of Theorem 4. By Lemma 2 this relation must have the form $A \equiv A_1 I = A_1 \sum I_j \mu_{jr}$. If $I = \tau_i I \tau_i^{-1}$, Lemma 1 implies that τ is Γ -dependent on the τ_j , $j = 1, \dots, h$. Hence $I \in H$ and $A = A_1 I \in G$.

The special case of this theorem in which $G = H$ consists of inner automorphisms is due to Artin and Whaples.⁸ We note also that if G is a finite

⁸ [1], p. 104.

group of automorphisms A_i such that each $A_i \neq 1$ is outer, then G is closed and of finite order. Hence Theorem 5 applies in this case too. The result thus obtained has been given in a previous paper of the author's.⁹ A more general result can be obtained by combining Theorem 5 with the following theorem:

THEOREM 6. *Let F be any finite group of automorphisms and let J be its subgroup of inner automorphisms. Then the closed group G generated by F has finite reduced order. If $J = \{\tau_1 \tau_1^{-1}, \dots, \tau_m \tau_m^{-1}\}$ then the reduced order of G is the product of the order of the set (τ_1, \dots, τ_m) over Γ by the index of J in F .*

Since J is closed under multiplication $\tau_i \tau_j = \tau_k \gamma_{ij}$, γ_{ij} in Γ . If A is any automorphism and I is the inner automorphism $\tau_i \tau_i^{-1}$ then $A^{-1}IA$ is the inner automorphism $(\tau_i A)_i (\tau_i A)_i^{-1}$. Hence if $A \in F$, $(\tau_i A)_i (\tau_i A)_i^{-1} \in J$. It follows that $\tau_i A = \tau_h \delta_h$, δ_h in Γ . Now let A denote the totality of linear combinations $\sum \tau_i \gamma_i$, γ_i in Γ . Since the product of any two τ 's is a multiple of a τ , A is a subalgebra. Clearly A has a finite basis and the dimensionality $(A : \Gamma)$ is the order of the set (τ_1, \dots, τ_m) over Γ . Since A has a finite basis, A is a division algebra. Let H denote the group of inner automorphisms determined by the elements of A . Evidently H is contained in the closed group G generated by F . If $A \in F$ and $I = \tau_i \tau_i^{-1} \in H$ where $\tau = \sum \tau_i \gamma_i$ then $A^{-1}IA$ is the inner automorphism determined by the element $\sum (\tau_i A) (\gamma_i A)$. Since $\tau_i A = \tau_h \delta_h$, δ_h in Γ and $\gamma_i A \in \Gamma$, $A^{-1}IA = I'$ is in H . It follows that the totality G' of automorphisms of the form AI , A in F and I in H is a group. For the product $(A_1 I_1)(A_2 I_2) = A_1 A_2 (A_2^{-1} I_1 A_2) I_2$ is in this set and $(AI)^{-1} = I^{-1} A^{-1} = A^{-1} (AI^{-1} A^{-1})$ is in the set. If AI is inner, so is A . Hence A is in J . Thus $AI \in H$. This shows that H is the group of inner automorphisms contained in G' . Since H is a group determined by the elements of an algebra, G' is closed. Hence $G' = G$ the smallest closed group containing F . Since the elements of G have the form AI , A in F , it is easy to see that the index of H in G is the same as that of J in F . It follows that G is of finite reduced order and that its reduced order is the product of $(A : \Gamma)$ by the index of J in F .

6. Let G be an arbitrary closed group of finite reduced order and let $\mathfrak{A} = GP_r$. We wish to establish a (1 — 1) correspondence between the closed subgroups K of G and the subrings \mathfrak{Q} of \mathfrak{A} that contain P_r .

We note first that if \mathfrak{Q} is any subring of \mathfrak{A} containing P_r , then

⁹ [3], p. 4.

$K = \mathfrak{Q} \circ G$ is a closed subgroup of G . Evidently K is closed under multiplication. If $A \in K$, A has an inverse. Hence A is not a zero-divisor in \mathfrak{Q} . Since \mathfrak{Q} satisfies the chain conditions this implies that $A^{-1} \in \mathfrak{Q}$. Hence K is a group. Now let $I_1 = \tau_{1l}\tau_{1r}^{-1}$ and $I_2 = \tau_{2l}\tau_{2r}^{-1} \in K$. Then $\tau_{3l} = I_1\tau_{1r} + I_2\tau_{2r} \in \mathfrak{Q}$ and if $\tau_3 \neq 0$, $I_3 = \tau_{3l}\tau_{3r}^{-1} \in \mathfrak{Q}$. Since G is closed $I_3 \in G$. Thus $I_3 \in K$. It follows that K is a closed group of automorphisms.

Next let K be any closed subgroup of G and let $\mathfrak{Q} = KP_r$. Clearly \mathfrak{Q} is a subring of \mathfrak{A} containing P_r . Let $K' = \mathfrak{Q} \circ G$. Then $K' \supseteq K$. On the other hand the elements of K' leave invariant every K -invariant. Hence, by Theorem 5, $K' \subseteq K$. Thus $K' = \mathfrak{Q} \circ G = K$.

Finally let \mathfrak{Q} be any ring between \mathfrak{A} and P_r . Let $K = \mathfrak{Q} \circ G$ and let $\mathfrak{Q}' = KP_r$. We wish to show that $\mathfrak{Q}' = \mathfrak{Q}$. Let $L = \mathfrak{Q} \circ H$. Then L is the subgroup of inner automorphisms in K . We may choose representatives A_1, A_2, \dots, A_u of the cosets of H in G such that A_1, \dots, A_t are representatives of the cosets of L in K . Also we may choose the inner automorphisms $I_1 = \tau_{1l}\tau_{1r}^{-1}, \dots, I_h = \tau_{hl}\tau_{hr}^{-1}$ in such a way that (τ_1, \dots, τ_h) is a basis for the algebra A associated with G while (τ_1, \dots, τ_f) is one for the algebra B associated with K . Now suppose that the element $V = \sum A_i I_j \mu_{jir} \in \mathfrak{Q}$. We wish to show that $V \in \mathfrak{Q}'$. Since $A_i I_i \in K$ if $i \leq t$, and $j \leq f$ it suffices to assume that $\mu_{ji} = 0$ for $j \leq f, i \leq t$ and to prove under this assumption that if $V \in \mathfrak{Q}$, $V = 0$.

Suppose then that there exists a $V \neq 0$ in \mathfrak{Q} such that if $V = \sum A_i I_j \mu_{jir}$ then $\mu_{ji} = 0$ for $j \leq f, i \leq t$. We may assume also that V is an element satisfying these conditions that has the fewest number of non-zero coefficients μ_{jir} . Then by the argument used to prove Lemma 2 we see that $V = A_i \sum_j I_j \mu_{jir} = A_i \sum_j \tau_{jl} \tau_{jr}^{-1} \mu_{jir}$. By the argument used to prove Lemma 1 we can see that we may assume that the elements $\gamma_j = \tau_j^{-1} \mu_{ji} \in \Gamma$. Hence $V = A_i \sum_j \tau_{jl} \gamma_j$ and since $V \neq 0$, $\tau = \sum \tau_j \gamma_j \neq 0$. Thus $V = A_i I_\tau$ where $I = \tau_l \tau_r^{-1}$. Evidently $A_i I \in K = \mathfrak{Q} \circ G$. Hence $i \leq t$ and τ is a linear combination of the τ_j with $j \leq f$. This contradicts the assumption on the form of V . We have therefore proved the following theorem:

THEOREM 7. *Let G be a closed group of automorphisms of finite reduced order and let $\mathfrak{A} = GP_r$. Then any subring \mathfrak{Q} of \mathfrak{A} containing P_r has the form KP_r where K is a closed subgroup of G . Moreover, distinct K 's give rise in this way to distinct rings \mathfrak{Q} .*

Let Φ be the division ring of G -invariants. Then we know that \mathfrak{A} is the complete ring of l. t. of P over Φ . We have also seen that there is a (1 — 1) correspondence between the rings \mathfrak{Q} between \mathfrak{A} and P_r and the division rings

between Φ and P . If we combine this (1—1) correspondence with the (1—1) correspondence between the rings \mathfrak{L} and the closed subgroups K of G we obtain the following result which includes the basic theorem of the ordinary Galois theory:

THEOREM 8. *Let P be a division ring and let G be a closed group of automorphisms in P of finite reduced order. To each closed subgroup K of G we associate the division subring $\Psi = P(K)$ of elements that are invariant under every $B \in K$ and to each division subring Ψ of P containing $\Phi = P(G)$ we associate the closed subgroup $K = G(\Psi)$ of automorphisms of G that leave the elements of Ψ invariant. Then these two correspondences are inverses of each other, and each is (1—1) between the closed subgroups of G and the division rings between P and Φ .*

7. It is an unsolved problem to determine conditions on a division subring Φ of P in order that there exist a closed group G of finite reduced order whose ring of invariants is Φ . If P is commutative we know that necessary and sufficient conditions are that P be finite, separable and normal over Φ .

Suppose now that $(P : \Gamma) < \infty$ and that Φ is a subfield of Γ such that $(\Gamma : \Phi) < \infty$ and let G be a closed group of automorphisms of P that has Φ as ring of invariants. Let H be the closed subgroup of G of automorphisms that leave the elements of Γ invariant. Now it is well known that an automorphism that produces the identity effect in Γ is inner. Hence H coincides with the subgroup of inner automorphisms of G . If A is the subalgebra associated with H we know that $(A : \Gamma) = (P : \Gamma)$. Hence $A = P$. Thus G includes all the inner automorphisms. The automorphisms A of G induce automorphisms \bar{A} in Γ and the correspondence A to \bar{A} is a homomorphism of G on a group \bar{G} of automorphisms in Γ . The kernel of this homomorphism is the set of automorphisms that leave the elements of Γ invariant. Hence the kernel is H and $\bar{G} \cong G/H$. Now Φ is the field of invariants of the group G acting in Γ . Hence Γ is finite, separable and normal over Φ .

Conversely suppose that P is finite over Γ and that Γ is finite, separable and normal over Φ . Assume, moreover, that every automorphism \bar{A} of the Galois group of Γ over Φ can be extended to an automorphism A in P . Let G be the totality of automorphisms of the form AI where I is inner. If $\bar{A}\bar{B} = \bar{C}$ and A, B, C , respectively, are the extensions of $\bar{A}, \bar{B}, \bar{C}$ then AB and C differ by an inner automorphism. Hence $AB = CI$. It follows easily that G is a group of automorphisms containing all of the inner automorphisms and that Φ is the ring of G -invariants.

Thus we have proved the following result which is in essence due to

Teichmüller:¹⁰ If Φ is a subfield of the center Γ of P and $(P : \Phi)$ is finite then necessary and sufficient conditions that there exist a closed group G of automorphisms whose set of invariants is Φ are 1) Γ is separable and normal over Φ and 2) every automorphism of the Galois group of Γ over Φ can be extended to an automorphism in P .

When these conditions hold we have seen that G is a group extension of the group H of inner automorphisms of P by the finite Galois group \tilde{G} of Γ over Φ . It is natural to seek conditions that this extension split in the sense that G contains a subgroup G' whose elements are representatives of the cosets of $G/H \cong \tilde{G}$. It is easy to see that this is equivalent to the requirement that P be a direct product $\Sigma \times \Gamma$ where Σ and Γ are algebras over Φ . Conditions for this have been given by Teichmüller.¹¹

THE JOHNS HOPKINS UNIVERSITY.

BIBLIOGRAPHY.

-
- [1] E. Artin and G. Whaples, "The theory of simple rings," *American Journal of Mathematics*, vol. 65 (1943), pp. 87-107.
 - [2] N. Jacobson, *The Theory of Rings*, Mathematical Surveys, vol. 2, New York, 1943.
 - [3] ———, "The fundamental theorem of the Galois Theory for quasi-fields," *Annals of Mathematics*, vol. 41 (1940), pp. 1-7.
 - [4] ———, "Structure theory of simple rings without finiteness assumptions," *Transactions of the American Mathematical Society*, vol. 57 (1945), pp. 228-245.
 - [5] E. Noether, "Nichtkommutative Algebra," *Mathematische Zeitschrift*, vol. 37 (1933), pp. 514-541.
 - [6] O. Teichmüller, "Über die sogenannte nichtkommutative Galoissche Theorie etc.," *Deutsche Mathematik*, vol. 5 (1940), 138-149.
- See also the review in the *Zentralblatt*, vol. 23, p. 198.

¹⁰ [6].

¹¹ [6].

THE DOUBLE CHAIN CONDITION IN CYCLIC OPERATOR GROUPS.*

By REINHOLD BAER.

In attempting to generalize Fitting's Lemma¹ the author found² that this was possible for abelian operator groups meeting the following requirements:

(a) the descending chain condition is satisfied by the admissible subgroups;

(b) if Z is a cyclic subgroup (i. e., if Z is generated by one element), then the ascending chain condition is satisfied by the admissible subgroups of Z .

Naturally one wonders whether condition (b) is a consequence of (a); and added strength is given to such a conjecture, if one remembers Hopkins' ³ Theorem to the effect that in rings possessing an identity the ascending chain condition for right-ideals is a consequence of the descending chain condition. However, it is possible to construct examples of cyclic groups where the descending, though not the ascending, chain condition is satisfied by the admissible subgroups.

In the light of Hopkins' Theorem, just quoted, it is only natural to assume that there will exist a large class of groups where cyclicity and descending chain condition imply the ascending chain condition; and it is the object of the present note to exhibit such classes of groups.

Let A be an abelian group where the composition is written as addition $a + b$; and assume that A admits the elements in the system M as operators (= right multipliers). The M -subgroup S of A is said to be of *length* $n = n(S)$, if the M -group S possesses a composition series⁴ of length n .

An M -subgroup J of A is termed *minimal*, if $J \neq 0$ and if J does not

* Received January 3, 1946. Presented to the American Mathematical Society December 26, 1946.

¹ See Jacobson (1), p. 9 for a statement of Fitting's Lemma. Various applications of this proposition may be found in Jacobson's book.

² Baer (2).

³ Apart from Hopkins' original paper, see Baer (1) for a proof, under weakened hypotheses, or Jacobson (1), p. 71, Theorem 29.

⁴ See e. g., Jacobson (1), p. 7.

possess any M -subgroups apart from 0 and J . This is equivalent to saying that minimal M -subgroups are the M -subgroups of length 1. The sum of all the minimal M -subgroups of A will be called the *socle*⁵ $S(A)$ of A . It is well known⁶ that double chain condition, ascending chain condition and descending chain condition for admissible subgroups are equivalent properties of the socle.

The ascending Loewy series $L(A, i)$ of A may now be defined⁷ inductively as follows: $L(A, 0) = 0$, $L(A, i + 1)$ is the uniquely determined M -subgroup of A satisfying: $L(A, i + 1)/L(A, i) = S(A/L(A, i))$. The subgroups $L(A, i)$ form an ascending chain of M -subgroups of A so that their (set-theoretical) join is an M -subgroup $L(A, \omega)$ of A . It is clear how to define the Loewy series for transfinite ordinals as well; but we have no need to do so.

LEMMA 1. $L(A, i)$ contains every M -subgroup of A whose length does not exceed i .

Proof. 0 is the only subgroup of length 0, proving our contention for $i = 0$. Thus we may assume the validity of our contention for $i - 1$ in order to prove it for i . If T is some M -subgroup of A such that $n(T) \leq i$, then T possesses a maximal M -subgroup T' so that $n(T) = n(T') + 1$. Since, therefore, $n(T') \leq i - 1$, it follows from our induction hypothesis that $T' \leq L(A, i - 1)$. Since T' is part of T also, this implies that $T' \leq T \cap L(A, i - 1) \leq T$. But T/T' is of length 1; and thus $T/[T \cap L(A, i - 1)]$ is of length 0 or 1. This latter group is isomorphic to $[T + L(A, i - 1)]/L(A, i - 1)$ so that this last group is either 0 or a minimal subgroup of $A/L(A, i - 1)$. It is therefore part of the socle of $A/L(A, i - 1)$ proving that $T + L(A, i - 1) \leq L(A, i)$. Hence $T \leq L(A, i)$ whenever $n(T) \leq i$, completing the proof.

COROLLARY 1. If the double chain condition is satisfied by the M -subgroups of $L(A, i)/L(A, i - 1)$ for every positive i , then the following properties of the M -group A imply each other:

- (i) $A = L(A, \omega)$.
- (ii) The cyclic M -subgroups⁸ of A are of finite length.

Proof. It is readily seen that every $L(A, i)$ is of finite length. If (i)

⁵ Following Remak. Note that the socle may be equal to 0.

⁶ See e.g. Jacobson (1), p. 14, Theorem 12.

⁷ Following Krull.

⁸ An M -group is termed cyclic, if it is generated, as an M -group, by one element.

is satisfied, then every cyclic M -subgroup C of A is part of some $L(A, i)$. But $L(A, i)$ is of finite length and thus every M -subgroup of $L(A, i)$ is of finite length. Hence C is of finite length, proving that (ii) is a consequence of (i).

If, conversely, (ii) is valid, then it follows from Lemma 1 that every element in A belongs to some $L(A, i)$; and thus (i) is a consequence of (ii).

COROLLARY 2. *The double chain condition is satisfied by the M -subgroups of A if, and only if,*

(a) *the double chain condition is satisfied by the M -subgroups of $L(A, i)/L(A, i-1)$ for every positive i ; and*

(b) *there exists an integer N such that every cyclic M -subgroup of A is of length not exceeding N .*

Proof. If the double chain condition is satisfied by the M -subgroups of A , then A has a definite finite length N which is an upper bound for the length of all its M -subgroups. This proves the necessity of condition (b); and the necessity of (a) is a consequence of the well known fact that subgroups and quotient groups of groups of finite length are of finite length too.

If, conversely, conditions (a) and (b) are satisfied, then we infer from Lemma 1 that $A = L(A, N)$ by (b); and this group is of finite length by (a).

Example 1. Denote by p a prime number and by A a direct sum of a countable infinity of cyclic groups of order p . Then A possesses a basis $b(1), b(2), \dots, b(i), b(i+1), \dots$. There exists an endomorphism $s(i)$ of A such that $b(j)s(i) = 0$ for $j \neq 2i+1$ and $b(2i+1)s(i) = b(2i)$; and there exists an endomorphism $t(i)$ of A such that $b(j)t(i) = 0$ for $j \neq 2i+2$ and $b(2i+2)t(i) = b(2i)$. The $s(i)$ and $t(i)$ for $0 < i$ form the system M .

It is readily seen that $L(A, i)$ for $0 < i$ is generated by the elements $b(1), b(3), \dots, b(2i-1), b(2i)$; and that $b(2i-1), b(2i)$ represent a basis of $L(A, i)/L(A, i-1)$ so that this group is of length 2. Thus the M -group A meets the requirement (a) of Corollary 2 in the strengthened form:

(a') *the groups $L(A, i)/L(A, i-1)$ for $0 < i$ are of bounded length.*

Since every element in A belongs to some $L(A, i)$, we see that the following form of condition (b) of Corollary 2 is satisfied by A :

(b') $A = L(A, \omega)$.

But it is clear that the ascending chain condition is not satisfied by the M -subgroups of A , since $L(A, i-1) < L(A, i)$ for $0 < i$. The descending chain condition is not satisfied by the M -subgroups of A either, since $b(2i-1)$ for $0 < i$ is not contained in the M -subgroup of A which is generated by the elements $b(2i), b(2i+1), \dots, b(2i+j), \dots$.

Note, furthermore, that (b') is just the condition (i) of Corollary 1 and that (a') implies the general hypothesis of Corollary 1. Thus the groups meeting all the requirements of Corollary 1 need not satisfy any of the chain conditions.

Following H. Prüfer we say that *the M -group A is of finite rank, if there exists an integer R with the following property:*

(R) *If U is a finite subset of A , then there exists an M -subgroup V of A which is generated by not more than R elements and contains U .*

Every finitely generated M -group is of finite rank through the converse is not true, as may be seen from the example of the groups of type p^∞ . Quotient groups of groups of finite rank are of finite rank too; but subgroups of groups of finite rank need not be a finite rank, as may be seen from the following

Example 2. If p is a prime number, A a direct sum of a countable infinity of cyclic groups of order p , and if $b(0), b(1), \dots, b(i), \dots$ is a basis of A , then there exists one, and only one, endomorphism $a(i)$ for $0 < i$ which maps $b(0)$ upon $b(i)$ and $b(j)$ for $0 < j$ upon 0. Denote by M the system consisting of the endomorphisms $a(i)$.

Clearly A is a cyclic M -group generated by $b(0)$. Denote by B the subgroup of A which is generated by the elements $b(i)$ for $0 < i$. Every subgroup of B is M -admissible; and it is clear that B is not of finite rank.

This Example 2 shows furthermore that we would have obtained a narrower class of groups, if we had substituted for condition (R) the following stricter condition.

(R') If the M -subgroup U of A is finitely generated, then it may be generated by R of its elements.

LEMMA 2. *The double chain condition is satisfied by the M -subgroups of A if, and only if,*

- (a) *A is of finite rank; and*
- (b) *the lengths of the cyclic M -subgroups of A are bounded.*

Proof. If the double chain condition is satisfied by the M -subgroups of A , then A is finitely generated, proving the necessity of (a); and the necessity of (b) is a consequence of Corollary 2.

Suppose, conversely, the validity of conditions (a) and (b). Denote by R an integer such that every finite subset of A is contained in a M -subgroup of A which is generated by not more than R elements; and denote by N an integer such that the length of every cyclic M -subgroup of A is at most N . If an M -subgroup of A is generated by not more than i elements, then it is clearly of a length not exceeding iN ; and the length of every finitely generated M -subgroup of A consequently does not exceed RN . Hence there exists an M -subgroup B of greatest length (note that M -subgroups of finite length are always finitely generated). If Z is a cyclic M -subgroup of A , then Z is of finite length and hence $B + Z$ is of finite length. This length cannot be greater than the length of B , proving (by the Jordan-Hölder Theorem) that $Z \leq B$. Hence $B = A$ so that A itself is of finite length, as we wished to show.

Again it is easy to show that neither of the conditions (a) and (b) can be omitted.

It is clear on reading the proofs of Lemmas 1 and 2 that the arguments used are purely lattice theoretical. It would be easy, though too lengthy for our purposes, to state these lemmas as theorems on lattices.

Example 3. Denote by p a prime number and let A be a direct sum of a countable infinity of cyclic groups of order p . Denote by $b, b(1), b(2), \dots, b(i), \dots$, for $0 < i$, a basis of A . Then there exists one and only one endomorphism $a(i)$ of A which maps b upon $b(i)$ and which maps every $b(j)$ upon 0; and there exists one and only one endomorphism $\alpha(i)$ of A which maps $b(i+1)$ upon $b(i)$ and which maps b as well as $b(j)$, for $j \neq i+1$, upon 0. Denote by M the system consisting of the $a(i)$ and $\alpha(i)$.

It is clear that A is a cyclic M -group, generated by b . Denote by $S(i)$ the subgroup of A which is generated by $b(1), \dots, b(i)$; and by $S(\omega)$ the subgroup of A which is generated by all the $b(i)$. These subgroups of A are readily seen to be M -admissible; and they satisfy:

$$0 < S(1) < \dots < S(i) < S(i+1) < \dots < S(\omega) < A.$$

Consider now some M -subgroup S of A . If S is not part of $S(\omega)$, then S contains an element of the form: $b + b'$ where b' is in $S(\omega)$. But then S contains $(b + b')a(i) = b(i)$ for every i . Hence b' and therefore b are in S , proving $S = A$, if S is not part of $S(\omega)$.

If the M -subgroup S of A contains an element that belongs to $S(i+1)$, but not to $S(i)$, then S contains an element of the form $b(i+1) + z$ for z in $S(i)$. Consequently S contains $(b(i+1) + z)\alpha(i) = b(i), b(i)\alpha(i-1) = b(i-1)$ etc., so that $S(i) \leq S$. Hence z is in S also which implies that $b(i+1)$ belongs to S . Consequently $S(i+1) \leq S$.

From the results of the last two paragraphs we deduce that $0, S(i), S(\omega)$ and A are the only M -subgroups of A . Thus we see that A is a cyclic M -group and that the descending, but not the ascending chain condition is satisfied by the M -subgroups of A .

If the abelian group A admits the elements in the system M as operators, then denote by M^* the ring of endomorphisms of the ordinary abelian group A which is generated by the identity and by those endomorphisms of A which are induced in A by elements in M . A subgroup of A is M -admissible if, and only if, it is M^* -admissible. Hence it does not involve any loss in generality, if we assume in the future that M itself is a ring containing an identity 1 which acts as an identity on A so that $x1 = x$ for x in A . Nothing will be gained, however, by assuming that the elements in M are endomorphisms of A , since endomorphism rings of A are, in general, not endomorphism rings of the M -subgroups of A . [For different endomorphisms of A may induce the same endomorphism in a subgroup.]

If Z is part of M , then it is possible to consider the M -group A at the same time as a Z -group; and it will be possible to gain information about the M -group A from information concerning the Z -group A , since every M -subgroup of A is at the same time a Z -group of A (though not conversely). From the remarks in the preceding paragraph it follows that we may assume, without loss of generality, that Z is a subring of M which contains the identity of the ring M ; and we shall make this hypothesis throughout.

If Z is a subring of M and if x is an element in the M -group A , then we denote by $Q(x, Z)$ the set of all the elements z in Z , satisfying $xz = 0$. Clearly $Q(x, Z)$ is a right-ideal in Z which may be called the Z -order of x .

LEMMA 3. If $A = gM$ is a cyclic M -group and if Z is a subring of M such that

$$(a) \quad MQ(g, Z) \leq Q(g, Z)M,$$

(b) the descending chain condition is satisfied by the Z -subgroups of gZ ; then the cyclic Z -subgroups of A are of bounded length.

Proof. From the definition of the Z -order and from condition (a) we infer that

$$gZ Q(g, Z) \leq gMQ(g, Z) \leq g Q(g, Z)M = 0.$$

Hence $ZQ(g, Z) \leq Q(g, Z)$ so that $Q(g, Z)$ is a two-sided ideal in \bar{Z} . Mapping the element gy , for y in Z , upon the coset $Q(g, Z) + y$ we obtain an isomorphism of the Z -group gZ upon the Z -group $Z/Q(g, Z)$; and this isomorphism induces, in particular, an isomorphism of the lattice of Z -subgroups of gZ upon the lattice of right-ideals in the ring $Z/Q(g, Z)$. It follows from (b) that the descending chain condition is satisfied by the right-ideals of the ring $Z/Q(g, Z)$, a ring with unit. Applying Hopkins' Theorem⁹ we see now that the double chain condition is satisfied by the right-ideals in $Z/Q(g, Z)$. Using the isomorphism between gZ and $Z/Q(g, Z)$ we conclude that

the cyclic Z -group gZ is of finite length N .

If x is an element in $A = gM$, then there exists an element m in M such that $x = gm$. Hence it follows from (a) that

$$xQ(g, Z) = gmQ(g, Z) \leq gMQ(g, Z) \leq gQ(g, Z)M = 0.$$

We deduce, therefore, from the definition of the Z -order that

$$Q(g, Z) \leq Q(x, Z) \text{ for every } x \text{ in } A.$$

As before we see that the Z -groups xZ and $Z/Q(x, Z)$ are essentially the same; and that, therefore, the lattices of Z -subgroups of xZ is isomorphic to the lattice of right-ideals between $Q(x, Z)$ and Z . From $Q(g, Z) \leq Q(x, Z)$ and the fact that the lattice of right-ideals between $Q(g, Z)$ and Z is of length N , we infer now that the lattices of right-ideals between $Q(x, Z)$ and Z is of a length not exceeding N . The Z -group xZ is, therefore, of length not greater than N ; and this completes the proof.

THEOREM 1. *The double chain condition is satisfied by the M -subgroups of the cyclic M -group $A = gM$, if there exists a subring Z of M with the following properties:*

- (i) $MQ(g, Z) \leq Q(g, Z)M$,
- (ii) *the descending chain condition is satisfied by the Z -subgroups of A .*

Proof. It is an immediate consequence of conditions (i) and (ii) and of Lemma 3 that the conditions (a) and (b) of Corollary 2 are satisfied by the Z -subgroups of A . It follows from Corollary 2 that the double chain condition is satisfied by the Z -subgroups of A . But, the system of M -subgroups

⁹ Which we stated in the introduction.

of A is part of the system of Z -subgroups of A , proving the validity of the double chain condition for the M -subgroups of A .

COROLLARY 3. *The double chain condition is satisfied by the M -subgroups of the cyclic M -group A , if either of the following conditions is satisfied by A and M :*

(1) *Every right-ideal in M is two-sided; and the descending chain condition is satisfied by the M -subgroups of A .*

(2) *If C is the center of M , then the descending chain condition is satisfied by the C -subgroups of A .*

This is an immediate consequence of Theorem 1.

The ring M shall be said to be of *finite rank over its subring Z* , if the addition group M_+ of M is a Z -group¹⁰ of finite rank.

LEMMA 4. *If $A = gM$ is a cyclic M -group, and if M is of finite rank over its subring Z , then A is a Z -group of finite rank.*

Proof. There exists an integer R with the following property:

If F is a finite subset of M , then there exist elements $m(1), \dots, m(k)$ in M such that $k \leq R$ and such that $F \leq \sum_{i=1}^k m(i)Z$.

Consider now a finite subset T of A . Then there exists a finite subset F of M such that $T = gF$. Hence there exist elements $m(1), \dots, m(k)$ such that $k \leq R$ and $F \leq \sum_{i=1}^k m(i)Z$. Clearly $T = gF \leq \sum_{i=1}^k [gm(i)]Z$, proving that the Z -group A is of finite rank.

THEOREM 2. *The double chain condition is satisfied by the M -subgroups of the cyclic M -group $A = gM$, if there exists a subring Z of M with the following properties:*

- (i) $MQ(g, Z) \leq Q(g, Z)M$;
- (ii) *the descending chain condition is satisfied by the Z -subgroups of gZ ;*
- (iii) M *is of finite rank over Z .*

Proof. It is a consequence of Lemmas 3 and 4 that conditions (a) and (b) of Lemma 2 are satisfied by the Z -subgroups of A . We infer from

¹⁰ The elements in Z act as right multipliers only for M_+ .

Lemma 2 the validity of the double chain condition for the Z -subgroups of A , implying the double chain condition for M -subgroups.

COROLLARY 4. *The double chain condition is satisfied by the M -subgroups of the cyclic M -group $A = gM$, if M is of finite rank over its center C and if the descending chain condition is satisfied by the C -subgroups of gC .*

This is an immediate consequence of Theorem 2.

UNIVERSITY OF ILLINOIS,
URBANA, ILLIOIS.

BIBLIOGRAPHY.

R. Baer.

- (1) "Radical Ideals," *American Journal of Mathematics*, vol. 64 (1943), pp. 537-568.
- (2) "Splitting Endomorphisms," *Transactions of the American Mathematical Society*.

Ch. Hopkins.

- (1) "Rings with minimal condition for left ideals," *Annals of Mathematics*, vol. 40 (1939), pp. 712-730.

N. Jacobson.

- (1) *The theory of rings*, Mathematical Surveys, No. 2, 1943.

RADICALS AND SUBDIRECT SUMS.*¹

By BAILEY BROWN and NEAL H. MCCOY.

1. Introduction. The radical N of a ring R , in which the right (left) ideals satisfy the descending chain condition, is usually defined as the join of all nil right (left) ideals in R . It is well known that N is a two-sided ideal in R and that the ring R/N is semi-simple, that is, has zero radical. Furthermore, a necessary and sufficient condition that R be semi-simple is that R be isomorphic to a full direct sum of a finite number of simple rings, each of which is a complete matrix ring over a division ring (Wedderburn-Artin theorems).

Various definitions have been proposed for the radical of an arbitrary ring,² most of which are in terms of nil ideals. Jacobson, however, has abandoned nilpotence as fundamental in the notion of radical;³ instead, his definition is in terms of a concept first introduced by Perlis.⁴ If N' is the Jacobson radical of R and $N' \neq R$, that is, R is not a radical ring, it turns out that N' is the intersection of a set of two-sided ideals B in R such that R/B is of a special type called primitive. From this, it follows that $N' = 0$ if and only if R is isomorphic to a subdirect sum of primitive rings. In the presence of the descending chain condition for right ideals, a primitive ring is necessarily simple, and a subdirect sum of a finite number of simple rings is actually a full direct sum of a finite number of these rings. Jacobson's results thus generalize the Wedderburn-Artin structure theorems mentioned above.

Our point of view, based somewhat on the form of Jacobson's results, is that the radical N of an arbitrary ring should be a two-sided ideal whose vanishing is a necessary and sufficient condition that R be isomorphic to a subdirect sum of rings of some particular type. Furthermore, if the descending chain condition holds for right ideals in R the radical should coincide with the classical one. We shall presently define a radical which generally differs from Jacobson's but which, in common with his, meets both these conditions.

* Received May 18, 1946.

¹ Presented to the Society, April 27, 1946.

² Köthe [1], Levitzki [1] and [2], Baer [1].

³ Jacobson [1].

⁴ Perlis [1].

First, however, we state his definition in terms which will conveniently lead to ours.

Let a be an arbitrary element of R , and consider the right ideal $\{ax - x\}$, where x runs over R .⁵ Now b is in the radical N' of R , as defined by Jacobson, if and only if $a \in \{ax - x\}$ for every element a in the right ideal generated by b . We shall henceforth refer to N' as the Jacobson radical of R .

Now let $G(a)$ be the two-sided ideal in R generated by the right ideal $\{ax - x\}$, that is,

$$G(a) = \{ax - x + \sum y_i a z_i - \sum y_i z_i\},$$

where x, y_i, z_i vary over R , the sums naturally being finite. We shall say that b is an element of the radical N of R if and only if $a \in G(a)$ for every element a of the two-sided ideal generated by b . Unless otherwise stated, this is henceforth the meaning we attach to the word "radical." We show that N is a two-sided ideal in R , that R/N has zero radical, and that the radical of a complete matrix ring R_n is N_n . *The vanishing of N is a necessary and sufficient condition that R be isomorphic to a subdirect sum of simple rings with unit element.*

If $b \in N'$, and a is an arbitrary element of the two-sided ideal generated by b , then, since N' is itself a two-sided ideal, it follows that $a \in \{ax - x\}$, so that $a \in G(a)$ and $b \in N$. Thus $N' \subseteq N$ and, since a primitive ring need not be a simple ring with unit element, in general, N' is properly contained in N . However, the definitions coincide if R is commutative, and hence for commutative rings, $N' = N$. We shall show that also $N' = N$ if the descending chain condition holds for right ideals in R . Thus our results furnish another generalization of one of the Wedderburn-Artin structure theorems and also, in common with Jacobson's, yield as special cases a number of known results on subdirect sums.⁶

Actually, we use a somewhat more general approach which may be of some interest in itself. In the proofs of some of the theorems about the radical, the only property of the $G(a)$ defined above which is actually used is that if $a \rightarrow \bar{a}$ is a homomorphism of R onto \bar{R} , then $G(\bar{a}) = \overline{G(a)}$. Accordingly, we consider an arbitrary mapping $a \rightarrow F(a)$, defined for all rings, of R into the set of two-sided ideals in R such that $F(\bar{a}) = \overline{F(a)}$ for every homomorphism $a \rightarrow \bar{a}$ of R onto a ring \bar{R} . If, then, in the definition of radical, we use $F(a)$ in place of $G(a)$, we get what we shall call the F -radical of R ; thus the radical

⁵ Jacobson uses $\{ax + x\}$, but we find it convenient to use this equivalent formulation.

⁶ Stone [1], McCoy [1], McCoy and Montgomery [1].

is the G -radical. It is shown that the F -radical N_F of R is a two-sided ideal in R and that R/N_F has zero F -radical. *The vanishing of N_F is a necessary and sufficient condition that R be isomorphic to a subdirect sum of subdirectly irreducible rings of zero F -radical.*⁷ Although the F -radical has a number of "radical-like" properties, it will not in general reduce to the ordinary radical in the presence of the descending chain condition. The concept is, therefore, not a true generalization of the radical, but rather a device leading to subdirect sum decompositions of various types, depending on the choice of the mapping F .

In view of these results, it becomes of some interest to characterize the subdirectly irreducible rings of zero F -radical. In the final section, we do this for several different mappings $a \rightarrow F(a)$, thus obtaining some other interesting special cases of the general subdirect sum decomposition theorem stated above.

Inasmuch as we shall be chiefly concerned with two-sided ideals the word *ideal* shall henceforth, unless otherwise stated, mean *two-sided ideal*.

2. Subdirect sum of rings. An element of the (full) direct sum S of rings R_i will be denoted by

$$a = (a_1, a_2, \dots),$$

where $a_i \in R_i$ and i , despite the notation employed, runs over an index set of arbitrary cardinal number. By a *subdirect sum* of the rings R_i is meant a subring T of S such that, for each i , the homomorphism $a \rightarrow a_i$ of S onto R_i maps T on all of R_i , inducing a homomorphism of T onto R_i . If a ring R is isomorphic to T , the product of the isomorphism and this homomorphism is a *natural* homomorphism of R onto R_i . *A ring R is isomorphic to a subdirect sum of rings R_i if and only if, for each i , R contains an ideal M_i with $R/M_i \cong R_i$, and the ideals M_i have zero intersection.*⁸

By a *subdirectly irreducible* ring⁷ is meant a ring R such that in every isomorphic representation of R as a subdirect sum of rings R_i , the natural homomorphism of R onto R_i is an isomorphism for some i . Thus, R is *subdirectly irreducible if and only if the intersection of all nonzero ideals in R is itself a nonzero ideal J* . This ideal J is a unique minimal ideal in R if R has more than one element and is vacuous otherwise.

⁷ The concept of subdirectly irreducible ring was introduced by Birkhoff [1]. See 2 below.

⁸ McCoy and Montgomery [1], McCoy [1].

3. The F -radical of a ring. Suppose that in each ring R there is assigned a mapping F of R into the set of ideals in R in such a way that if $a \rightarrow \bar{a}$ defines a homomorphism of R onto a ring \bar{R} , and $\overline{F(a)}$ is the image in \bar{R} of the ideal $F(a)$ in R , then $F(\bar{a}) = \overline{F(a)}$. An element a of R shall be F -regular if and only if $a \in F(a)$, and an ideal in R shall be F -regular if and only if each of its elements is F -regular.

DEFINITION. The F -radical N_F of R is the set of elements b of R which generate F -regular ideals (b) in R .

If $N_F = R$, R is an F -radical ring.

If $a \in F(a)$, then $\bar{a} \in \overline{F(a)} = F(\bar{a})$. It follows that F -regularity of elements, and therefore of ideals, is preserved under homomorphism, and that the F -radical of R goes into the F -radical of \bar{R} .

The following theorem characterizes the F -radical:^o

THEOREM 1. The F -radical N_F of a ring R is the intersection $\cap M$ of the ideals M in R such that R/M is subdirectly irreducible and has zero F -radical.

We show first that $\cap M \subseteq N_F$. If $b \notin N_F$, then for some a in (b) , $a \notin F(a)$. By Zorn's Maximum Principle, there is an ideal M containing $F(a)$ but not containing a , such that every ideal in R which contains M as proper subset does contain a . If \bar{a} is the residue class to which a belongs modulo M , it follows that every nonzero ideal in R/M contains $\bar{a} \neq 0$, so R/M is subdirectly irreducible. But since $F(\bar{a}) = \overline{F(a)} = \bar{0}$, $\bar{a} \notin F(\bar{a})$. Thus every nonzero ideal in R/M contains an element which is not F -regular, so that R/M has zero F -radical. Since $a \notin M$, it follows that $b \notin M$, and $b \notin \cap M$.

Conversely, $N_F \subseteq \cap M$. For if $b \in N_F$ and M is any ideal of the specified type, then \bar{b} is in the F -radical of R/M since F -regularity is preserved under homomorphism. Since R/M has zero F -radical, $\bar{b} = \bar{0}$, $b \in M$. Thus $b \in \cap M$.

COROLLARY 1. The F -radical of R is an ideal in R .

COROLLARY 2. A ring R is an F -radical ring if and only if R itself is the only ideal M in R such that R/M is subdirectly irreducible of zero F -radical.

COROLLARY 3. The F -radical of R is the join of all F -regular ideals in R .

This last statement follows from the fact that, by definition, N_F consists of the elements of R which generate F -regular ideals. Since N_F is an ideal it is clearly an F -regular ideal which contains every F -regular ideal.

^o Cf. Birkhoff [1]; Jacobson [1], p. 311.

THEOREM 2. *The ring R/N_F has zero F -radical.*

Let \bar{b} denote an element of the F -radical of R/N_F , and M any ideal in R such that R/M is subdirectly irreducible and has zero F -radical. Since $R/M \cong (R/N_F)/(M/N_F)$, it follows from Theorem 1 that $\bar{b} \in M/N_F$. Hence $b \in M$, $b \in \text{IIM} = N_F$, $\bar{b} = \bar{0}$.

THEOREM 3. *A subdirect sum of rings R_i of zero F -radical has zero F -radical.*

If a is in the F -radical of R and $a \rightarrow a_i$ is the natural homomorphism of R onto R_i , then, for each i , a_i is in the F -radical of R_i . Hence $a_i = 0$ and $a = 0$.

Almost immediate is

THEOREM 4. *A ring has zero F -radical if and only if it is isomorphic to a subdirect sum of subdirectly irreducible rings of zero F -radical.*

The necessity of the condition follows from Theorem 1 and the results of 2; the sufficiency, from Theorem 3.

If, for example, the mapping F is defined in each ring R by $F(a) = 0$ for every a in R , zero is the only F -regular element of R and every ring has zero F -radical. Theorem 4 then reduces to the ring case of Birkhoff's theorem on algebras with finitary operations.⁷

We now prove

THEOREM 5. *A subdirectly irreducible ring R has zero F -radical if and only if the minimal ideal J contains an element $a \neq 0$ such that $F(a) = 0$.*

If R has zero F -radical and $j \neq 0$ is an element of J , some element a of $(j) = J$ is not F -regular, that is, $a \notin F(a)$. This element a can not be zero and is contained in every nonzero ideal in R . Hence we must have $F(a) = 0$. Conversely, the F -radical of R does not contain J , for, by hypothesis, J contains an element a which is not F -regular. Hence the F -radical is the zero ideal.

In the next two sections, we make a detailed study of the radical, that is, the case in which the general mapping $a \rightarrow F(a)$ is specialized to $a \rightarrow G(a)$, where $G(a)$ is as defined in the introduction. In 6, further illustrations and applications of this general theory will be given.

4. The radical of a ring.

We now repeat the

DEFINITION. The *radical* N of a ring R is the set of all elements b of R such that for every element a of (b) ,

$$a \in G(a) = \{ax - x + \sum y_i a z_i - \sum y_i z_i\}.$$

A ring with zero radical will be said to be *semi-simple*; a ring which coincides with its radical is a *radical ring*. If a ring has more than one element and a unit element 1, it is clear that $1 \notin G(1) = 0$, so that such a ring can not be a radical ring.

THEOREM 6. *A subdirectly irreducible ring is semi-simple if and only if it is a simple ring with unit element.*

The theorem is trivial for one-element rings.¹⁰ If the simple ring S has more than one element and a unit element 1, then $1 \in J$, $G(1) = 0$, and Theorem 5 shows that S is semi-simple. Conversely, let S be a subdirectly irreducible ring of more than one element. If S is semi-simple, Theorem 5 states that its minimal ideal J contains an element $e \neq 0$ for which $G(e) = 0$. Hence $ex = x$ for all x in S and $S \subseteq J$, that is, S is simple and has a left unit e . The proof is completed by establishing the following lemma:¹¹

LEMMA. *If e is a left unit of a simple ring S , then e is the unit element of S .*

It is given that $ex = x$ for all x in S . The set $E = \{xe - x; x \in S\}$ is a two-sided ideal in S . If $E = S$, then $e = ye - y$ for some y in S , and right multiplication by e yields $e = ye - ye = 0$. Hence $x = ex = 0$ for all x , and S is a one-element ring. If S has more than one element, then $E = 0$, and $x = ex = xe$ for all x in S .

COROLLARY. *A simple ring is semi-simple if it has a unit element, otherwise it is a radical ring.*

The following characterization of the radical follows at once from Theorems 1 and 6:

THEOREM 7. *The radical of a ring R is the intersection of all ideals M in R such that R/M is a simple ring with unit element.*

COROLLARY 1. *If R is not a radical ring, the radical of R is the intersection of all maximal ideals M in R such that R/M has a unit element.*

¹⁰ In order to avoid any possible confusion, we explicitly point out that, according to our usage, a one-element ring is a simple ring with unit element.

¹¹ This was pointed out to us by R. E. Johnson.

COROLLARY 2. *If R is not a one-element ring and has a unit element, the radical of R is the intersection of all maximal ideals in R .*

Theorems 4 and 6 yield at once the following important result:

THEOREM 8. *A ring R is semi-simple if and only if it is isomorphic to a subdirect sum of simple rings with unit element.*

This is, for general rings, a generalization of one of the Wedderburn-Artin theorems for semi-simple rings in which the right ideals satisfy the descending chain condition. We shall show that it actually yields that theorem if merely the descending chain condition for two-sided ideals is assumed.

THEOREM 9. *In the presence of the descending chain condition for two-sided ideals, a semi-simple ring R is isomorphic to the full direct sum of a finite number of simple rings with unit element.*

Since R has zero radical, there exists a set of ideals M in R such that R/M is simple with unit element, and $\Pi M = 0$. Since the descending chain condition is assumed, there is a finite subset $\{M_i\}$, ($i = 1, 2, \dots, k$), of these ideals with $\Pi M_i = 0$, and which, furthermore, is minimal in the sense that the intersection of any $k - 1$ ideals of the subset is not zero. If $a \in R$ and a_i is the image in R/M_i of the element a under the natural homomorphism of R onto R/M_i , it is clear that the correspondence

$$(1) \quad a \rightarrow (a_1, a_2, \dots, a_k)$$

defines an isomorphism of R onto a subdirect sum of the simple rings R/M_i . However, it is easy to show as follows that this subdirect sum is actually the full direct sum. From the minimality of $\{M_i\}$ it is clear that there is an element c of R in $\prod_{i=2}^k M_i$, but not in M_1 . For this c , the correspondence (1) yields $c \rightarrow (c_1, 0, \dots, 0)$. Since c_1 is a nonzero element of the simple ring R/M_1 , it follows that $(c_1) = R/M_1$. Hence, under the isomorphism (1), every element of the form $(r_1, 0, \dots, 0)$, $r_1 \in R/M_1$, appears as the image of some element of R . A similar result holds if M_1 is replaced by M_j , ($j = 2, \dots, k$). Thus, finally, every element of the direct sum of the rings R/M_i appears as the image of an element of R under the isomorphism (1). This shows that R is isomorphic to the direct sum of the rings R/M_i , ($i = 1, 2, \dots, k$).

It was pointed out in the introduction that the Jacobson radical N' is contained in the radical N . We now prove

THEOREM 10. *If the right ideals of R satisfy the descending chain condition, the radical N of R coincides with the Jacobson radical N' of R .*

In view of the remark just made, it is enough to prove that $N \subseteq N'$. In R/N' , the descending chain condition holds for right ideals and the Jacobson radical is zero. This implies³ that R/N' is isomorphic to the direct sum of a finite number of simple rings, each of which has a unit element because of the chain condition. Thus R/N' has zero radical by Theorem 8. Hence if $b \in N$, then \bar{b} is in the radical of R/N' , so $\bar{b} = \bar{0}$ and $b \in N'$. Thus $N \subseteq N'$, as required.

If R has more than one element and a unit element, Jacobson has shown that N' is the intersection of all maximal right (left) ideals in R . The second Corollary to Theorem 7, together with the result just established, shows that if R has a unit element and the descending chain condition holds for right ideals in R , the intersection of all maximal right (left) ideals coincides with the intersection of all maximal two-sided ideals.

We shall conclude this section with some further miscellaneous results. First, we remark that if a is a nilpotent element of a ring R , then $a \in G(a)$. For if $a^n = 0$, $n > 1$, and $x = -\sum_{i=1}^{n-1} a^i$, it is easily verified³ that $a = ax - x$. This establishes

THEOREM 11. *If all elements of (b) are nilpotent, then b is in the radical of R .*

We now pass to the proof of

THEOREM 12. *If A is an ideal in R , the radical of the ring A is contained in the radical N of R .*

Let M be any ideal in R for which R/M is a simple ring with unit element. Then $A/(A \cap M) \cong (A, M)/M$, which, as an ideal in the ring R/M , is itself a simple ring with unit element. Thus, by Theorem 7, the radical of the ring A is contained in $A \cap M$, and therefore in $N = \Pi M$.

COROLLARY. *Any ideal in a semi-simple ring is semi-simple.*

5. The radical of a matrix ring. The ring of all matrices of order n with elements in a ring R will be denoted, as usual, by R_n . The purpose of this section is to prove

THEOREM 13. *The radical of R_n is N_n .*

If M is an ideal in R , it is clear that $M \rightarrow M_n$ is a one-one mapping of the set of ideals in R onto a subset of the ideals in R_n , and that $M \subseteq M'$ if and only if $M_n \subseteq M'_n$.

LEMMA 1. Let L be an ideal in R_n and a the element in the (i, j) position of a matrix A in the ideal L . If x and y are elements of R , then L contains a matrix with xay in the (p, q) position and zeros elsewhere.

If $x \in R$, let X_{ij} denote the matrix with x in the (i, j) position and zeros elsewhere. Then $X_{pi}AY_{jq}$ is the required matrix in L .

LEMMA 2. S is a simple ring with unit element if and only if S_n is a simple ring with unit element.

If S is simple with unit element and L a nonzero ideal in S_n , let $a \neq 0$ be an element of some matrix in L . In S , the ideal $(a) = \{\sum x_i a y_i\} = S$, and hence by Lemma 1, if $s \in S$, there exists in L a matrix with s appearing in any given position and zeros elsewhere. Adding matrices of this type, it is clear that $L = S_n$ and it is obvious that S_n has a unit element. Conversely, the mapping $M \rightarrow M_n$ establishes that S_n is not simple if S is not, and it is well known that S_n has no unit element if S does not.

The theorem will follow readily when we have proved

LEMMA 3. If R/M is a simple ring with unit element, then R_n/M_n is a simple ring with unit element. Conversely, if L is an ideal in R_n such that R_n/L is a simple ring with unit element, then the set M of all elements of R which appear as elements in the matrices of L is an ideal in R , $L = M_n$ and R/M is a simple ring with unit element.

The first assertion follows from Lemma 2 and the readily proved fact that $R_n/M_n \cong (R/M)_n$.

For the converse, let M^* be the set of elements of R which appear in the $(1, 1)$ position in matrices of L . Clearly M^* is an ideal in R and $M^* \subseteq M$. We show that $M \subseteq M^*$ and hence that $M = M^*$.

Since, by hypothesis, R_n/L has a unit element, R_n contains a matrix $U = (u_{ij})$ which is the unit element of R_n modulo L . Thus

$$(2) \quad UXU \equiv X \pmod{L}$$

for every matrix X in R_n . If, in particular, we choose X to be a matrix with an arbitrary element r of R in the $(1, 1)$ position and zeros elsewhere, and consider the elements in the $(1, 1)$ position of the two members of (2), we see that

$$(3) \quad u_{11}ru_{11} \equiv r \pmod{M^*}$$

for every element r of R . If $b \in M$, there is a matrix in L containing b in some position. Lemma 1 then shows that there exists a matrix in L con-

taining $u_{11}bu_{11}$ in the $(1, 1)$ position. Hence $u_{11}bu_{11} \in M^*$ and it follows from (3) that $b \in M^*$. Thus $M \subseteq M^*$, and M is an ideal in R .

We show next that $L = M_n$. It is evident that $L \subseteq M_n$, so it suffices to prove that $M_n \subseteq L$. If $c \in M$, then $UC_{hk}U$ can be considered as a sum of matrices of the type described in Lemma 1, hence $UC_{hk}U \in L$. From (2), it follows that $C_{hk} \in L$. But any matrix of M_n can be expressed as a sum of matrices of the form C_{hk} . Thus $M_n \subseteq L$ and therefore $L = M_n$.

Finally, since $(R/M)_n \cong R_n/M_n = R_n/L$, which is a simple ring with unit element, it follows from Lemma 2 that R/M is a simple ring with unit element. This completes the proof of Lemma 3.

It is apparent from Lemma 3 and the properties of the mapping $M \rightarrow M_n$ that the ideals M in R for which R/M is a simple ring with unit element correspond one-to-one with the ideals L in R_n for which R_n/L is a simple ring with unit element. Furthermore, it is clear that intersections of corresponding ideals also correspond. It follows from Theorem 7 that the radicals correspond, and thus the radical of R_n is N_n .

COROLLARY. *If R is a radical ring, R_n is a radical ring.*

6. Some other special cases of the general theory. In this section, we return to give some further illustrations and miscellaneous results of the general theory of 3. We recall that the mapping $a \rightarrow F(a)$ of R into the set of ideals in R is assumed to have the property that if $a \rightarrow \bar{a}$ is a homomorphism of R onto \bar{R} , then $F(\bar{a}) = F(a)$.

In addition to the important special case, $a \rightarrow G(a)$, discussed in detail in the preceding two sections, we may list the following expressions for $F(a)$, all of which are ideals and meet the above stated condition. The list is by no means exhaustive but suggests some of the possibilities. In the following, α is an arbitrary integer:

$$F_a(a) = \{ax - \alpha x + ya - \alpha y + \sum r_i a s_i - \alpha \sum r_i s_i\},$$

$$F_{a^1}(a) = \{ax - \alpha x + \sum r_i a s_i - \alpha \sum r_i s_i\},$$

$$F_{a^r}(a) = \{xa - \alpha x + \sum r_i a s_i - \alpha \sum r_i s_i\},$$

$$H_a(a) = \{\sum r_i a s_i - \alpha \sum r_i s_i\},$$

$$G^*(a) = G(-a^2) = \{a^2 x + x + \sum r_i a^2 s_i + \sum r_i s_i\}.$$

In these expressions; x, y, r_i, s_i vary over the elements of R and the sums are finite.

It will be noted that the $G(a)$, used in the definition of the radical, is just $F_{a^1}(a)$.

Obviously, $F_a(a) = 0$ if and only if $ax = xa = \alpha x$ for all x in R , that is, a is an α -fier of R .¹² Thus, as a consequence of Theorem 5, we obtain

THEOREM 14. *A subdirectly irreducible ring R has zero F_a -radical if and only if the minimal ideal J contains a nonzero α -fier of R .*

This theorem together with Theorems 1, 6 and 7 show that the F_1 -radical of a ring coincides with the radical. In fact, the F_1 -radical is also just the radical.

Consider next the mapping defined by $a \rightarrow G^*(a)$. Here $G^*(a) = 0$ if and only if $-a^2$ is a left unit of R . Theorems 4 and 5, together with the argument used in the proof of Theorem 6, establish at once the following result:

THEOREM 15. *A ring has zero G^* -radical if and only if it is isomorphic to a subdirect sum of simple rings, each with a unit element 1 and an element whose square is -1 .*

The principle used in constructing $G^*(a)$ can obviously be generalized by using, in place of $-a^2$, any polynomial in a with integral coefficients and zero constant term. This process may also be applied to the other mappings as well as to G .

If R has a unit element, $F_a = F_{a'} = F_{a''} = H_a$, but in general they are different. We discuss, in some detail, the cases of H_a for $\alpha = 1$ and 0 respectively.

It is clear that $H_1(a) = 0$ if and only if

$$(4) \quad xay - xy = 0$$

for all x and y in R . Thus, by Theorem 5, a subdirectly irreducible ring R has zero H_1 -radical if and only if its minimal ideal J contains a nonzero element a for which (4) holds. We now consider two cases.

We suppose first that the subdirectly irreducible ring R has no right or left annihilator except zero. That is, $zR = 0$ implies $z = 0$, and $Rz = 0$ implies $z = 0$. From (4), we see that $(xa - x)R = 0$ and $R(ax - x) = 0$ for every x in R . Hence $ax = xa = x$, and a is the unit element of R . But $a \in J$, so R is a simple ring with unit element.

Suppose, on the other hand, that R has a nonzero element c such that $cR = 0$ or $Rc = 0$, and, for definiteness, assume that $cR = 0$. It follows that (c) consists of all elements of the form $nc + rc$, where n is an integer

¹² Brown and McCoy [1].

and $r \in R$. But $a \in (c)$ since $J \subseteq (c)$, hence $aR = 0$. From (4), we have that $xy = 0$ for all x, y in R , that is, that $R^2 = 0$.¹³

Conversely, in either case, a subdirectly irreducible ring has zero H_1 -radical. Hence from Theorem 4 we obtain

THEOREM 16. *A ring has zero H_1 -radical if and only if it is isomorphic to a subdirect sum of rings, each of which is either a simple ring with unit element or a subdirectly irreducible ring whose square is zero.*

Finally, we consider the case of H_α in which $\alpha = 0$. Here $H_0(a) = 0$ if and only if

$$(5) \quad xay = 0$$

for all x and y in R . As in the preceding case, we only need to characterize those subdirectly irreducible rings R which have a nonzero element a of the minimal ideal J such that (5) is satisfied. Let us assume for the moment that $aR \neq 0$, hence that there exists an element z of R such that $az \neq 0$. Now the ideal (az) must contain a , hence in view of (5) we have

$$a = naz + azr,$$

where n is an integer and $r \in R$. From this, again using (5), it follows that $Ra = 0$. Hence either $Ra = 0$ or $aR = 0$. Since $(a) = J$, it follows at once that either $RJ = 0$ or $JR = 0$. It is clear that if this condition is satisfied, the subdirectly irreducible ring R has zero H_0 -radical. Theorem 4 now establishes

THEOREM 17. *A ring has zero H_0 -radical if and only if it is isomorphic to a subdirect sum of subdirectly irreducible rings R_i with the minimal ideal J_i such that either $R_iJ_i = 0$ or $J_iR_i = 0$.*

It is an easy consequence of Theorem 14 that in the decompositions induced by the vanishing of the F_0 -radical, the rings R_i satisfy $J_iR_i = R_iJ_i = 0$. For the F_0^l -radical (F_0^r -radical), they satisfy $J_iR_i = 0$ ($R_iJ_i = 0$).

AMHERST COLLEGE.
SMITH COLLEGE.

¹³ A detailed characterization of subdirectly irreducible rings R with $R^2 = 0$ would take us too far afield. However, it is not difficult to show that the additive group of such a ring is primary of rank one, as defined by Prüfer. Hence such a group is either of the type p^n or of the type p^∞ (Prüfer [1]).

BIBLIOGRAPHY.

R. Baer:

- [1] "Radical ideals," *American Journal of Mathematics*, vol. 65 (1943), pp. 537-568.

G. Birkhoff:

- [1] "Subdirect unions in universal algebra," *Bulletin of the American Mathematical Society*, vol. 50 (1944), pp. 764-768.

B. Brown and N. H. McCoy:

- [1] "Rings with unit element which contain a given ring," *Duke Mathematical Journal*, vol. 13 (1946), pp. 9-20.

N. Jacobson:

- [1] "The radical and semi-simplicity for arbitrary rings," *American Journal of Mathematics*, vol. 67 (1945), pp. 300-320.

G. Köthe:

- [1] "Die Struktur der Ringe, deren Restlassenring nach dem Radikal vollständig reduzibel ist," *Mathematische Zeitschrift*, vol. 32 (1930), pp. 161-186.

J. Levitzki:

- [1] "On the radical of a general ring," *Bulletin of the American Mathematical Society*, vol. 49 (1943), pp. 462-466.
[2] "Semi-nilpotent ideals," *Duke Mathematical Journal*, vol. 10 (1943), pp. 553-556.

N. H. McCoy:

- [1] "Subrings of infinite direct sums," *Duke Mathematical Journal*, vol. 4 (1938), pp. 486-494.
[2] "Subdirectly irreducible commutative rings," *Duke Mathematical Journal*, vol. 12 (1945), pp. 381-387.

N. H. McCoy and D. Montgomery:

- [1] "A representation of generalized Boolean rings," *Duke Mathematical Journal*, vol. 3 (1937), pp. 455-459.

S. Perlis:

- [1] "A characterization of the radical of an algebra," *Bulletin of the American Mathematical Society*, vol. 48 (1942), pp. 128-132.

H. Prüfer:

- [1] "Untersuchungen über die Zerlegbarkeit der abzählbaren primären Abelschen Gruppen," *Mathematische Zeitschrift*, vol. 17 (1923), pp. 35-61.

M. H. Stone:

- [1] "The theory of representations for Boolean algebras," *Transactions of the American Mathematical Society*, vol. 40 (1936), pp. 37-111.

ON CONGRUENCES AND CONJUGATE NETS.*

By V. G. GROVE.

1. Introduction. In this paper we study conjugate nets from a metric point of view. The equations of the sustaining surface, assumed to be non-developable, are written in terms of arbitrary parameters. The formulas convenient for the study of the net are written in terms of these parameters and certain normalized components of vectors along the tangents to the curves of the net. An invariant of a curve of the net, which we have called the *asymptotic curvature* of the curve, has been found which is related to isothermally conjugate nets in a manner similar to the relation between the geodesic curvature of the curves of an orthogonal net to isothermic nets. Conditions that a congruence of lines protruding from the surface be conjugate to the sustaining surface, and to the net are found. Some attention is paid to pencils of conjugate nets.

Let the parametric equations of the surface S be $x^i = x^i(u^1, u^2)$ and let $X^i(u^1, u^2)$ be the direction cosines of the normal to S at the point x with coordinates (x^1, x^2, x^3) . These functions satisfy the differential equations

$$(1.1) \quad x^i_{,\alpha\beta} = d_{\alpha\beta}X^i, \quad X^i_{,\alpha} = m_{\alpha}^{\rho}x^i_{,\rho}, \quad m_{\alpha}^{\beta} = -d_{\alpha\rho}g^{\rho\beta}$$

the comma denoting covariant differentiation with respect to the metric tensor $g_{\alpha\beta}$. Let $du^1/du^2 = U^1/U^2$, $du^1/du^2 = V^1/V^2$ be the differential equations of the conjugate net N . Then $d_{\rho\sigma}U^{\rho}V^{\sigma} = 0$. We shall call the curve of N whose tangent vector is $U^{\alpha}(V^{\alpha})$ the U -curve (V -curve) with similar names for the tangents to these curves.

Since only the ratios $U^1 : U^2$, $V^1 : V^2$ are material, these components may be normalized so that

$$(1.2) \quad d_{\rho\sigma}U^{\rho}U^{\sigma} = e_1, \quad d_{\rho\sigma}V^{\rho}V^{\sigma} = e_2$$

wherein e_1, e_2 are ± 1 according as the radii of normal curvature $R_{(u)}$, $R_{(v)}$ in the directions of the U - and V -tangents are positive or negative. From (1.2) and the definitions of these curvatures we have

$$R_{(u)} = e_1 g_{\rho\sigma} U^{\rho} U^{\sigma}, \quad R_{(v)} = e_2 g_{\rho\sigma} V^{\rho} V^{\sigma}.$$

* Received July 9, 1946.

Since S is non-developable we may define the functions $d^{a\beta}$ by the expression $d^{a\rho}d_{\rho\beta} = \delta_{\beta}^a$. Let

$$U_a = d_{a\rho}U^\rho, \quad V_a = d_{a\rho}V^\rho.$$

The following relations exist between U^a, V^a :

$$(1.3) \quad \begin{aligned} e_1 U^a U^\beta + e_2 V^a V^\beta &= d^{a\beta}, & e_1 U_a U_\beta + e_2 V_a V_\beta &= d_{a\beta}, \\ U^a V^\beta - V^a U^\beta &= \epsilon^{a\beta}, & V^a &= \epsilon^{a\rho} U_\rho, & U^a &= \epsilon^{a\rho} V_\rho \end{aligned}$$

wherein

$$\begin{aligned} \epsilon^{11} = \epsilon^{22} &= 0, & \epsilon^{12} = -\epsilon^{21} &= (ed)^{-\frac{1}{2}}, & e &= e_1 e_2 \\ \epsilon_{11} = \epsilon_{22} &= 0, & \epsilon_{12} = -\epsilon_{21} &= (ed)^{\frac{1}{2}}. \end{aligned}$$

2. S referred to N , and the asymptotic net. Let us now write the differential equations (1.1) of S in terms of the conjugate parameters of N . To this end put

$$(2.1) \quad x^i_1 = f U^\rho x^i_{,\rho}, \quad x^i_2 = g V^\rho x^i_{,\rho}, \quad F = \log f, \quad G = \log g.$$

wherein f and g are to be chosen so that $(x^i_1)_2 = (x^i_2)_1$. We find from (2.1) that

$$x^i_{,a} = e_1 U_a x^i_1 / f + e_2 V_a x^i_2 / g.$$

and, since $(x^i_1)_2 = (x^i_2)_1$

$$(2.2) \quad V^\rho F_{,\rho} = e_1 U_\rho W^\rho, \quad U^\rho G_{,\rho} = -e_2 V_\rho W^\rho,$$

wherein

$$W^a = U^\rho V^a_{,\rho} - V^\rho U^a_{,\rho}.$$

The functions x^i, X^i satisfy the equations

$$(2.3) \quad x^i_{a\beta} = l^p_{a\beta} x^i_{,p} + D_{a\beta} X^i, \quad X^i_a = M^p_a x^i_{,p},$$

wherein

$$(2.4) \quad \begin{aligned} l^1_{11} &= e_1 f (e_1 U^\rho F_{,\rho} + U_\rho U^{\rho,\sigma} U^\sigma), & l^1_{22} &= e_1 g^2 U_\rho V^{\rho,\sigma} V^\sigma / f, \\ l^2_{11} &= e_2 f^2 V_\rho U^{\rho,\sigma} U^\sigma / g, & l^2_{22} &= e_2 g (e_2 V^\rho G_{,\rho} + V_\rho V^{\rho,\sigma} V^\sigma), \\ l^1_{12} &= e_1 g U_\rho V^{\rho,\sigma} U^\sigma, & l^2_{12} &= e_2 f V_\rho U^{\rho,\sigma} V^\sigma, \\ D_{11} &= e_1 f^2, & D_{12} &= 0, & D_{22} &= e_2 g^2, \\ M^1_1 &= -e_1 h_{\rho\sigma} U^\rho U^\sigma, & M^2_1 &= -e_2 f h_{\rho\sigma} U^\rho V^\sigma / g, \\ M^2_1 &= -e_1 g h_{\rho\sigma} U^\rho V^\sigma / f, & M^2_2 &= -e_2 h_{\rho\sigma} V^\rho V^\sigma, \end{aligned}$$

$h_{a\beta}$ being the metric tensor of the spherical representation of S .

One may show readily that, under the conditions (2.2), the expressions

$$du = U_\rho du^\rho / f, \quad dv = V_\rho dv^\rho / g$$

are exact, and hence (2.3) are the Gauss differential equations of S in the parameters u, v and in which $x^i_1 = \partial x^i / \partial u$, etc.

As an illustration of the use one may make of (2.3), we recall that the focal point on the U -tangent is given by the formula

$$\rho^i_{(u)} = x^i - x^i_1 / l^2_{12} = x^i - e_2 U^\rho x^i_{,\rho} / (V_\sigma U^\sigma{}_{,\lambda} V^\lambda).$$

Similarly the focal point on the V -tangent is given by

$$(2.5) \quad \rho^i_{(v)} = x^i - e_1 V^\rho x^i_{,\rho} / (U_\sigma V^\sigma{}_{,\lambda} U^\lambda).$$

Moreover the direction numbers λ^i of the axis of N are given by the formula

$$\begin{aligned} \lambda^i &= e_2 l^1_{22} x^i_1 / g^2 + e_1 l^2_{11} x^i_2 / f^2 + X^i \\ &= e [U_\rho V^\rho{}_{,\sigma} V^\sigma U^\lambda + V_\rho U^\rho{}_{,\sigma} U^\sigma V^\lambda] x^i_{,\lambda} + X^i. \end{aligned}$$

Again N is isothermally conjugate if and only if $(\log D_{11}/D_{22})_{12} = 0$. Using (2.1), (2.2) and (2.3) this condition reduces to the vanishing of the invariant I defined by

$$(2.6) \quad I = e_1 (U_\rho W^\rho)_{,\sigma} U^\sigma + e_2 (V_\rho W^\rho)_{,\sigma} V^\sigma.$$

Define the vectors A^a, B^a by the formulas

$$(2.7) \quad A^a = U^a + \sqrt{-e} V^a, \quad B^a = U^a - \sqrt{-e} V^a.$$

One finds readily that $d_{\rho\sigma} A^\rho A^\sigma = 0$, $d_{\rho\sigma} B^\rho B^\sigma = 0$, hence these vectors are tangent vectors of the asymptotic curves. We shall speak of these tangents as the A - and B -tangents.

Defining x^i_1, x^i_2 by the expressions

$$x^i_1 = m A^\rho x^i_{,\rho}, \quad x^i_2 = n B^\rho x^i_{,\rho}, \quad M = \log m, \quad N = \log n;$$

and demanding that $(x^i_1)_2 = (x^i_2)_1$ we find that m and n must satisfy the conditions

$$B^\rho M_{,\rho} = -e_1 \sqrt{-e} B_\rho W^\rho, \quad A^\rho N_{,\rho} = e_1 \sqrt{-e} A_\rho W^\rho.$$

It follows that x^i, X^i are solutions of equations of the form (2.3) with coefficients defined by

$$\begin{aligned} l^1_{11} &= \frac{1}{2} e_1 m A^\rho (2e_1 M_{,\rho} + B_\sigma A^\sigma{}_{,\rho}), & l^1_{22} &= \frac{1}{2} n^2 B_\rho B^\rho{}_{,\sigma} B^\sigma / m, \\ l^2_{11} &= \frac{1}{2} e_1 m^2 A_\rho A^\rho{}_{,\sigma} A^\sigma, & l^2_{22} &= \frac{1}{2} e_1 n B^\rho (2e_1 N_{,\rho} + A_\sigma B^\sigma{}_{,\rho}), \\ (2.8) \quad l^1_{12} &= \frac{1}{2} e_1 n B_\rho B^\rho{}_{,\sigma} A^\sigma, & l^2_{12} &= \frac{1}{2} e_1 m A_\rho A^\rho{}_{,\sigma} B^\sigma, \\ M^1_1 &= -\frac{1}{2} e_1 h_{\rho\sigma} A^\rho B^\sigma, & M^2_1 &= -\frac{1}{2} e_1 m h_{\rho\sigma} A^\rho A^\sigma / n, \\ M^1_2 &= -\frac{1}{2} e_1 n h_{\rho\sigma} B^\rho B^\sigma / m, & M^2_2 &= -\frac{1}{2} e_1 h_{\rho\sigma} A^\rho B^\sigma, \\ D_{11} &= D_{22} = 0, & D_{12} &= 2mn e_1. \end{aligned}$$

As an application, we observe that S is ruled if $R_1 R_2 = 0$ and a quadric if $R_1 = R_2 = 0$ where

$$(2.9) \quad R_1 = e_1 A_\rho A^\rho{}_{,\sigma} A^\sigma, \quad R_2 = e_2 B_\rho B^\rho{}_{,\sigma} B^\sigma.$$

Using (2.7) we may write (2.9) in the form

$$R_1 = \Phi - \sqrt{-e} \Psi, \quad R_2 = \Phi + \sqrt{-e} \Psi,$$

wherein

$$\begin{aligned} \Phi &= e_1 U_\rho U^\rho{}_{,\sigma} U^\sigma - e_2 (U_\rho V^\rho{}_{,\sigma} V^\sigma + V_\rho U^\rho{}_{,\sigma} V^\sigma + V_\rho V^\rho{}_{,\sigma} U^\sigma), \\ \Psi &= e_2 V_\rho V^\rho{}_{,\sigma} V^\sigma - e_1 (V_\rho U^\rho{}_{,\sigma} U^\sigma + U_\rho V^\rho{}_{,\sigma} U^\sigma + U_\rho U^\rho{}_{,\sigma} V^\sigma). \end{aligned}$$

Hence S is a quadric if and only if $\Phi = \Psi = 0$.

As a second application of this section, it is well known that the polar reciprocal l_2 of the line l_1 joining the points

$$r^{i(1)} = x^i - e_1 x^i{}_{,1} / (bm), \quad r^{i(2)} = x^i - e_1 x^i{}_{,2} / (an)$$

has direction numbers

$$x^i{}_{,12} - e_1 a n x^i{}_{,1} - e_1 b m x^i{}_{,2}.$$

Using (2.8) the direction numbers of l_2 assume the form

$$\lambda^i = \frac{1}{4} [(B_\rho B^\rho{}_{,\sigma} A^\sigma - 2a) A^\lambda + (A_\rho A^\rho{}_{,\sigma} B^\sigma - 2b) B^\lambda] x^i{}_{,\lambda} + X^i.$$

It follows that the reciprocal of the normal to S joins the points $r^{i(1)}$, $r^{i(2)}$ defined by

$$\begin{aligned} r^{i(1)} &= x^i - 2e_1 A^\rho x^i{}_{,\rho} / (A_\sigma B^\sigma{}_{,\lambda} B^\lambda), \\ r^{i(2)} &= x^i - 2e_1 B^\rho x^i{}_{,\rho} / (B_\sigma A^\sigma{}_{,\lambda} A^\lambda). \end{aligned}$$

3. The asymptotic curvature of a curve. Denoting covariant differentiation with respect to the tensor $d_{\alpha\beta}$ by a semicolon, we find from (1.2) that

$$d_{\rho\sigma} U^\rho U^\sigma{}_{;\lambda} U^\lambda = 0.$$

Hence the vector $U^\alpha{}_{;\rho} U^\rho$ is conjugate to the U -tangent, and we may write

$$(3.1) \quad U^\alpha{}_{;\rho} U^\rho = k_{(u)} V^\alpha.$$

Using the third of (1.3) we find that

$$(3.2) \quad k_{(u)} = \epsilon_{\rho\sigma} U^\rho U^\sigma{}_{;\lambda} U^\lambda.$$

We shall call the function $k_{(u)}$ defined by (3.1) or (3.2) *the asymptotic curvature of the U -curve*.

Similarly the asymptotic curvature of the V -curve is given by the expression

$$k_{(v)} = \epsilon_{\rho\sigma} V^\rho V^\sigma ;_\lambda V^\lambda.$$

An alternate form of (3.2) may be found by multiplying both members of (3.1) by V_α and summing on α . We find that

$$(3.3) \quad k_{(u)} = e_2 V_\rho U^\rho ;_\sigma U^\sigma.$$

From (1.2) we find that $U_\rho U^\rho ;_\sigma V^\sigma = 0$. Hence (3.3) may be written in the form

$$(3.4) \quad k_{(u)} = e_2 U^\rho ;_\sigma (V_\rho U^\sigma - U_\rho V^\sigma) = e_2 d_{\rho\lambda} \epsilon^{\sigma\lambda} U^\rho ;_\sigma = e \epsilon^{\rho\sigma} U_{\sigma,\rho}.$$

Since

$$U_\rho W^\rho = U_\rho (U^\sigma V^\rho ;_\sigma - V^\sigma U^\rho ;_\sigma) = (U^\rho V^\sigma - V^\rho U^\sigma) U_{\rho,\sigma},$$

we have other expressions for $k_{(u)}$, namely

$$(3.5) \quad k_{(u)} = -e_2 U_\rho W^\rho = -e V^\rho F_{,\rho}.$$

Similarly

$$(3.6) \quad k_{(v)} = e_1 \epsilon^{\rho\sigma} V_{\sigma,\rho} = e_1 V_\rho W^\rho = -e U^\rho G_{,\rho}.$$

If the parametric curves on S are conjugate and if these curves are chosen as the U - and V -curves, then we may write

$$\begin{aligned} U^1 &= (e_1 d_{11})^{-\frac{1}{2}}, & U^2 &= 0, & U_1 &= (e_1 d_{11})^{\frac{1}{2}}, & U_2 &= 0, \\ V^1 &= 0, & V^2 &= (e_2 d_{22})^{-\frac{1}{2}}, & V_1 &= 0, & V_2 &= (e_2 d_{22})^{\frac{1}{2}}. \end{aligned}$$

Then $k_{(u)}$ and $k_{(v)}$ may be written in the forms

$$(3.7) \quad k_{(u)} = -e_2 (ed)^{-\frac{1}{2}} \partial / \partial u^2 (e_1 d_{11})^{\frac{1}{2}}, \quad k_{(v)} = -e_1 (ed)^{-\frac{1}{2}} \partial / \partial u^1 (e_2 d_{22})^{\frac{1}{2}}.$$

These latter forms of the asymptotic curvatures of the curves of a conjugate net show their resemblance to the geodesic curvature of orthogonal curves.

Comparing (3.5) and (3.6) with (2.6) we find that *if the asymptotic curvatures of the curves of a conjugate net are constant along the corresponding curves of the net, the net is isothermally conjugate.*

The curve whose differential equation is $du^1/du^2 = C^1/C^2$ is an extremal¹ of the integral

$$I = \int \sqrt{ed_{\rho\sigma} du^\rho du^\sigma}$$

if the components C^α satisfy the equation

$$\epsilon_{\rho\sigma} C^\rho C^\sigma ;_\lambda C^\lambda = 0.$$

It follows from (3.2) that a non-asymptotic curve is an extremal of the integral I only if the asymptotic curvature of the curve vanishes at each of its points.

Now define the vectors \bar{U}^a , \bar{V}^a by the formulas

$$(3.8) \quad \begin{aligned} \bar{U}^a &= U^a \cos \theta + V^a \sin \theta, \\ \bar{V}^a &= -U^a \sin \theta + V^a \cos \theta, \end{aligned}$$

if $e_1 = e_2$, $e = 1$, that is, if $K > 0$; and by the formulas

$$(3.9) \quad \begin{aligned} \bar{U}^a &= U^a \cosh \theta + V^a \sinh \theta, \\ \bar{V}^a &= U^a \sinh \theta + V^a \cosh \theta, \end{aligned}$$

if $e_1 = -e_2$, $e = 1$, that is, if $K < 0$. In either case $d_{\rho\sigma}\bar{U}^\rho\bar{U}^\sigma = e_1$, $d_{\rho\sigma}\bar{V}^\rho\bar{V}^\sigma = e_2$, and hence these vectors are normalized. One may readily show that the \bar{U} - and \bar{V} -tangents are conjugate, and that the curves defined by them form a pencil of conjugate nets if and only if $\theta = \text{const.}$ Denoting the asymptotic curvatures of the \bar{U} -curves and \bar{V} -curves by $\bar{k}_{(u)}$ and $\bar{k}_{(v)}$ we find from (3.4) and (3.6) that

$$\begin{aligned} \bar{k}_{(u)} &= k_{(u)} \cos \theta - k_{(v)} \sin \theta, \\ \bar{k}_{(v)} &= k_{(u)} \sin \theta + k_{(v)} \cos \theta, \quad K > 0; \end{aligned}$$

and

$$\begin{aligned} \bar{k}_{(u)} &= k_{(u)} \cosh \theta + k_{(v)} \sinh \theta, \\ \bar{k}_{(v)} &= k_{(u)} \sinh \theta + k_{(v)} \cosh \theta, \quad K < 0. \end{aligned}$$

For these two cases therefore we observe that

$$\begin{aligned} \bar{k}_{(u)}^2 + \bar{k}_{(v)}^2 &= k_{(u)}^2 + k_{(v)}^2, \quad K > 0, \\ \bar{k}_{(u)}^2 - \bar{k}_{(v)}^2 &= k_{(u)}^2 - k_{(v)}^2, \quad K < 0. \end{aligned}$$

In either case we observe that if the curves of a conjugate net are the extremals of the integral I , then all curves of the pencil determined by the net are also extremals of I ; moreover the given net is isothermally conjugate.

If we denote the transform of the invariant I defined by (2.6) under the transformations (3.8) with $\theta = \text{const.}$ by \bar{I} , we find that $\bar{I} = I$. We therefore corroborate Wilczynski's theorem³ that if one net of a pencil of conjugate nets is isothermally conjugate, all nets of the pencil are isothermally conjugate.

If $K > 0$, the associate conjugate net of the given net N is given by (3.8) with $\theta = \pi/4$, and if $K < 0$ by (3.9) with $\theta = \pi i/4$. Suppose N is such that its associate net form the lines of curvature. Then from

$g_{\rho\sigma}\bar{U}^\rho\bar{V}^\sigma = 0$, $\theta = \pi/4$, we find that the radii of normal curvature of the curves of N are given by

$$R_{(u)} = R_{(v)} = \frac{1}{2}(R_1 + R_2)$$

R_1, R_2 being the principal radii of normal curvature. *The net whose associate net is the lines of curvature is then the net of mean curvature.* The same result follows from (3.9), it being observed that the given net is then imaginary, the lines of curvature being real. The radii of normal curvature in the directions of the tangents to the curves of any net (3.8) whose basis net is the net of mean curvature are given by

$$\bar{R}_{(u)} = R_1 \cos^2 \left(\theta - \frac{\pi}{4} \right) + R_2 \sin^2 \left(\theta - \frac{\pi}{4} \right),$$

$$\bar{R}_{(v)} = R_1 \sin^2 \left(\theta - \frac{\pi}{4} \right) + R_2 \cos^2 \left(\theta - \frac{\pi}{4} \right).$$

From (2.4) we observe that the U -curve is a geodesic if $l^2_{11} = 0$, that is if

$$V_\rho U^\rho{}_{;\sigma} U^\sigma = 0.$$

Moreover its asymptotic curvature is zero if and only if

$$(3.10) \quad V_\rho U^\rho{}_{;\sigma} U^\sigma = 0.$$

If $L^a{}_{\beta\gamma}$ and $\Lambda^a{}_{\beta\gamma}$ are respectively the Christoffel symbols formed from the tensors $g_{a\beta}$, $d_{a\beta}$, one finds readily that

$$\Lambda^a{}_{\beta\gamma} = L^a{}_{\beta\gamma} + T^a{}_{\beta\gamma}$$

wherein

$$(3.11) \quad T^a{}_{\beta\gamma} = \frac{1}{2} d^{a\rho} d_{\beta\gamma, \rho}.$$

If the U -curve is a geodesic, equation (3.10) may be written in the form

$$T^\rho{}_{\sigma\lambda} V_\rho U^\sigma U^\lambda = 0$$

which by (3.11) may be written in the form

$$(3.12) \quad U_\rho V^\rho{}_{;\sigma} U^\sigma = 0.$$

Comparing (3.12) with (2.5) we observe that *if the U -curve is a geodesic, it has zero asymptotic curvature if and only if the ray of the net is parallel to the V -tangent.*

The meridians on a surface of revolution are known to be geodesics. We shall find the condition that their asymptotic curvature vanish. If the equations of S are written in the form

$$x^1 = u^1 \cos u^2, \quad x^2 = u^1 \sin u^2, \quad x^3 = \phi(u^1),$$

then

$$d_{11} = \phi''/(1 + (\phi')^2)^{\frac{1}{2}}, \quad d_{12} = 0, \quad d_{22} = u^1 \phi'/(1 + (\phi')^2)^{\frac{1}{2}}, \quad \phi' = d\phi/du^1.$$

Hence from (3.7) the asymptotic curvature of the parallels vanishes; the asymptotic curvature of the meridians vanishes if and only if

$$\partial/\partial u^1 \log \{(u^1 \phi')^2/1 + (\phi')^2\} = 0.$$

Integrating we find that

$$\phi = c \cosh^{-1} u^1/c.$$

Hence the only surfaces of revolution whose meridians have zero asymptotic curvatures are the catenoids; the parallels and meridians of this surface therefore form an isothermally conjugate net, a well known fact.

4. Congruences conjugate to S and N . Let there be given a congruence λ of lines l protruding from S at x . The components of the vector along l may be written in the form

$$(4.1) \quad \lambda^i = p(\theta^{\rho} x^i_{,\rho} + X^i), \quad p \neq 0, \quad P = \log p.$$

From (4.1) we find

$$(4.2) \quad \lambda^i_{,\sigma} = \mu^{\rho}_{\sigma} x^i_{,\rho} + (P_{,\sigma} + \theta_{\sigma}) \lambda^i,$$

wherein

$$\mu^{\rho}_{\sigma} = p(\theta^{\rho}_{,\sigma} + m^{\rho}_{\sigma} - \theta_{\sigma} \theta^{\rho}).$$

The curves on S corresponding to the developables of λ are the integral curves of the differential equation

$$\epsilon_{\rho\lambda} \mu^{\rho}_{\sigma} du^{\lambda} du^{\sigma} = 0.$$

It follows that λ is conjugate to S if and only if

$$(4.3) \quad \epsilon_{\rho\lambda} \mu^{\rho}_{\sigma} d^{\lambda\sigma} = 0.$$

Using the fact that $\epsilon^{a\beta}$ is skew symmetric, and that $d_{a\beta,\gamma} - d_{a\gamma,\beta} = 0$, condition (4.3) may be reduced¹ to

$$(4.4) \quad \theta_{a,\beta} - \theta_{\beta,a} = 0.$$

Condition (4.4) is equivalent to the condition that the differential equation $\theta_{\rho} du^{\rho} = 0$ is an exact differential equation. Let therefore $\omega(u^1, u^2) = \text{const.}$ be an arbitrary one-parameter family of curves whose

differential equation is $\omega_{,\rho} d\omega^\rho = 0$. A point y on the conjugate to the curve C of the family through x has coordinates defined by the expression

$$y^i = x^i + p d^{\rho\sigma} \omega_{,\sigma} x^i_{,\rho}.$$

Erect at y a perpendicular of length cp ($c = \text{const.}$) to the tangent plane to S , determining a point z . The direction numbers of the line l joining x to z are of the form

$$\lambda^i = p(d^{\rho\sigma} \omega_{,\sigma} x^i_{,\rho} + cX^i).$$

The most general congruence λ conjugate to S is generated by this line l .

A special case gives rise to an interesting class of congruences conjugate to S . Suppose in (4.1) that $d_{\rho\sigma} \theta^\rho \theta^\sigma = e_1$. The condition (4.4) that λ be conjugate to S , is equivalent to saying that the θ -curve has zero asymptotic curvature. Moreover this condition is invariant under the transformation $\theta_a = c\tilde{\theta}_a$ ($c = \text{const.}$). This class of congruences may be constructed as follows: *Select an arbitrary one parameter family of extremals of the integral I ; at the intersections of the tangent lines to these curves with the Dupin indicatrix, erect perpendiculars of constant length to the tangent planes to S . The lines l determined by the points x of S and the terminal points of these perpendiculars generate a congruence conjugate to S .*

From (4.2) one finds that the conditions² that the developables of λ correspond to a parametric net are

$$\mu^\alpha_\beta = 0, \quad \alpha \neq \beta.$$

Multiply (4.2) by fU^σ and then by gV^σ and sum on σ . In the notation of (2.1), the resulting equations may be written in the form

$$\lambda^i_a = v^\rho_a x^i_{,\rho} + A_a \lambda^i$$

wherein

$$v_1^2 = e_2 f \mu^\rho_\sigma U^\sigma V_\rho / g, \quad v_2^1 = e_1 g \mu^\rho_\sigma V^\sigma U_\rho / f,$$

the remaining coefficients not being material for our purposes. *The developables of λ therefore correspond to the U - and V -curves if and only if*

$$(4.5) \quad \mu^\rho_\sigma U^\sigma V_\rho = 0, \quad \mu^\rho_\sigma V^\sigma U_\rho = 0.$$

Conditions (4.5) are not independent; in fact the vanishing of their sum is implied by (4.4). If in the first of (4.5) we let $\theta^\rho = -d^{\rho\lambda}(\log \omega)_{,\lambda}$ that equation may be written in the equivalent forms

$$U^\sigma V^\lambda (\omega_{,\sigma\lambda} - 2T^\rho_{\sigma\lambda}\omega_{,\rho} - h_{\sigma\lambda}\omega) = 0,$$

$$U^\sigma V_\rho [(d^\rho\lambda\omega_{,\lambda})_{,\sigma} + d_{\sigma\lambda}g^\rho\lambda\omega] = 0.$$

This latter form is equivalent to Springer's result.²

5. Normal congruences. There is a striking similarity between the theory of congruences conjugate to a surface S , and that of congruences normal to a second surface S' . In (4.1) choose p so that $\lambda^i\lambda_i = 1$, and let $\phi^\alpha = p\theta^\alpha$. Then (4.1) assumes the form

$$\lambda^i = \phi^\rho x^\rho_{,i} + pX^i, \quad \lambda^i\lambda_i = 1 = p^2 + g_{\rho\sigma}\phi^\rho\phi^\sigma.$$

We may write the parametric equations of any transversal surface S' of λ in the form

$$y^i = x^i + q\lambda^i.$$

Then λ is the normal congruence of S' if $\lambda^iy^\alpha_{,a} = 0$ for $\alpha = 1, 2$. These conditions are equivalent to the conditions

$$g_{\alpha\rho}\phi^\rho + q_{,a} = 0.$$

Hence λ is normal to S' if and only if

$$(5.1) \quad (g_{\alpha\rho}\phi^\rho)_{,a} - (g_{\beta\rho}\phi^\rho)_{,a} = 0.$$

Let $\phi_a = g_{\alpha\rho}\phi^\rho$; then (5.1) becomes²

$$(5.2) \quad \phi_{a,\beta} - \phi_{\beta,a} = 0.$$

Let $\omega(u^1, u^2) = c$ be a one-parameter family of curves whose differential equation is $\omega_{,\rho}du^\rho = 0$. A point y on the tangent to S perpendicular to the tangent to the curve C of this family through x has coordinates

$$y^i = x^i + cg^{\rho\sigma}\omega_{,\sigma}x^\rho_{,i},$$

wherein c is a constant such that $c^2g^{\rho\sigma}\omega_{,\rho}\omega_{,\sigma} < 1$. At y erect a perpendicular to the tangent plane to S at x intersecting the unit sphere with center at x in the points z_1, z_2 . The line l determined by x and z_1 (or z_2) generates the most general normal congruence.

By analogy to the method in the latter part of 4, consider a vector ϕ^α of constant length c , that is let

$$g_{\rho\sigma}\phi^\rho\phi^\sigma = c^2.$$

We find that

$$g_{\rho\sigma}\phi^\rho\phi^\sigma_{,\lambda}\phi^\lambda = 0.$$

The vector $\phi^a{}_{,\rho}\phi^\rho$ is therefore perpendicular to ϕ^a . Then

$$\phi^a{}_{,\rho}\phi^\rho = c^2 k_g \psi^a, \quad g_{\rho\sigma}\psi^\rho\psi^\sigma = 1, \quad g_{\rho\sigma}\phi^\rho\psi^\sigma = 0$$

where k_g is the geodesic curvature of the ϕ -curve, k_g being given by

$$k_g = \sqrt{eK} \epsilon^{\rho\sigma} \phi_{\rho,\sigma}, \quad \phi_a = g_{a\rho} \phi^\rho.$$

Hence the condition (5.2), for a vector of constant length is equivalent to the vanishing of k_g . Our results may be stated as follows: *Select a one parameter family of geodesics on S , lay off points on these tangents at a constant distance from the points of contact; erect perpendiculars to the tangent planes of S of constant length from these points. The lines l determined by x and the terminal points of these segments generate a congruence normal to some surface S' .*

MICHIGAN STATE COLLEGE.

REFERENCES.

1. V. G. Grove, "A general theory of surfaces and conjugate nets," *Transactions of the American Mathematical Society*, vol. 57 (1945), pp. 105-122.
2. C. E. Springer, "Rectilinear congruences whose developables intersect a surface in its lines of curvature," *Bulletin of the American Mathematical Society*, vol. 51 (1945), pp. 990-996.
3. E. J. Wilczynski, "Geometrical significance of isothermal conjugacy," *American Journal of Mathematics*, vol. 42 (1920), pp. 211-221.

THE ELLIPTIC MODULAR FUNCTION AND A CLASS
OF ANALYTIC FUNCTIONS FIRST
CONSIDERED BY HURWITZ.*

By ZEEV NEHARI.

Introduction. In spite of its wide applicability in various branches of the theory of functions, the elliptic modular function is often used with a certain hesitation. This is mainly due to the fact that its application presupposes familiarity with a comparatively intricate formalism, in particular when the determination of numerical constants is involved. In fact, the endeavor to avoid the elliptic modular function has given rise to an extensive mathematical literature aiming at proving certain theorems in an "elementary" way, the word "elementary" being used here as a synonym for "without making use of the elliptic modular function." As an impressive example, Picard's theorem on integral functions might be quoted.

The difficulties which beset the numerical treatment of the elliptic modular function go essentially back to the fact that, on the one hand, the formalism of this function can only be developed with the help of the Jacobian elliptic functions while, on the other hand, what is needed in the applications are the conformal mapping properties of the modular function, and the connection between these two different aspects of the modular function has to be established through the medium of the theory of Schwarz' differential parameter or by a very detailed study of the periodic properties of the Jacobian elliptic functions.

The object of the first part of this paper is to show how those properties of the elliptic modular function which are required for the applications may be derived in a simple way by the exclusive use of elementary principles of the theory of conformal representation. It will be shown that once the "modular surface" is defined, the functional equation

$$(1) \quad J(z) = 4\sqrt{J(z^2)}/[1 + \sqrt{J(z^2)}]^2$$

satisfied by the elliptic modular function can be obtained almost by inspection. This functional equation—which is a special case of Landen's transformation of the Jacobian elliptic functions—has no other solution except the elliptic

* Received May 3, 1946.

modular function, if the further conditions $J(0) = 0$, $J'(0) > 0$ and $J(z)$ regular for $|z| < 1$, are stipulated, as then the Taylor coefficients of $J(z)$ can be successively computed with the help of (1). It will then be shown that the infinite product representing the square of the modulus of the Jacobian elliptic functions (taken as a function of the period ratio) likewise satisfies equation (1) and the additional conditions. Equation (1) having only one such solution, this implies the identity of $J(z)$ and the square of the Jacobian modulus and yields, at the same time, the convenient product expansion of the modular function.

It will further be shown that, apart from (1), $J(z)$ also satisfies some other functional relations which together amount to the fact that $J(e^{\pi i x})$ is an automorphic function of x , i. e., it is invariant with respect to a certain group of linear transformations of x .

In the second part of this paper, a number of theorems on functions $f(z) = a_1 z + a_2 z^2 + \dots$, regular for $|z| < 1$ and not vanishing there except at $z = 0$, are proved with the help of the elliptic modular function.

PART I.

1. The elliptic modular function $w = M(z)$ (the relationship between $M(z)$ and the function $J(z)$ referred to in the Introduction will appear later) may be defined as the analytic function effecting the conformal representation of a curvilinear triangle with zero angles in the z -plane on the half-plane $J_m\{w\} > 0$, the sides of the triangle being circular arcs orthogonal to the circle $|z| = 1$; the vertices of the triangle (which necessarily are situated on $|z| = 1$) may be made to correspond to the points $w = 0, 1, \infty$. By the well-known procedure of inverting the triangle along its sides, then again inverting the resulting figure along its sides etc. etc., the whole circle $|z| < 1$ will eventually be covered by an infinity of triangles. By Schwarz' symmetry principle, all these triangles are mapped by $w = M(z)$ on half-planes $J_m\{w\} > 0$ or $J_m\{w\} < 0$ which are connected with their neighboring half-planes along the stretch $0 < w < 1$ and the rays $-\infty < w < 0$ and $1 < w < \infty$, respectively. The full circle $|z| < 1$ is mapped by $w = M(z)$ on a Riemann surface, say R_1 , which covers the w -plane an infinity of times, with the exception of the points $w = 0, 1, \infty$ which are isolated boundary points of R_1 .

In many applications it is more convenient to use a Riemann surface slightly different from R_1 . This surface, say R_2 , differs from R_1 by the fact that while $w = 0, 1, \infty$ are all outside R_1 , one sheet of R_2 covers the point

$w = 0$. The function mapping $|z| < 1$ on R_2 will be denoted by $w = J(z)$. In virtue of Riemann's fundamental theorem, $J(z)$ is uniquely determined by the further conditions $J(0) = 0$ and $J'(0) > 0$.

Both $M(z)$ and $J(z)$ are called 'elliptic modular functions'; indeed they are connected by a simple relation. It is seen without difficulty that the function

$$v = u(z) = \beta \cdot e^{-\alpha[(1+\gamma z)/(1-\gamma z)]} \quad [\alpha > 0, |\beta| = |\gamma| = 1]$$

maps $|z| < 1$ on a surface which covers the circle $|v| < 1$ an infinity of times and has inside $|v| = 1$ the only boundary point $v = 0$, which moreover is a boundary point for all sheets of the surface. Therefore, if we form the function $J[u(z)]$, we see that the point $w = 0$ is "removed" from R_2 , R_2 thereby being transformed into R_1 . Hence

$$M(z) = J[\beta \cdot e^{-\alpha[(1+\gamma z)/(1-\gamma z)]}].$$

2. We shall now derive the fundamental functional equation of $J(z)$. To this end we consider the function $w = \sqrt{J(z^2)}$ which, as is easily seen, maps $|z| < 1$ on a surface whose only boundary points are $w = 0, \pm 1, \infty$, all of which are outside the surface, apart from $w = 0$ which is covered by one sheet only (for $z = 0$). As in the case of R_2 , there are no inner branch-points to this surface. We now consider the function

$$w^* = p(z) = 4\sqrt{J(z^2)}/[1 + \sqrt{J(z^2)}]^2.$$

The critical points $w = -1, 0, 1, \infty$ are transformed by $w^* = 4w/(1+w)^2$ into $w^* = \infty, 0, 1, 0$ respectively. Hence, $w^* = p(z)$ takes the value $w^* = 0$ only for $z = 0$ and leaves out the values $w^* = 1, \infty$. Since $dw^*/dw = 4(1-w)/(1+w)^3$ vanishes only for $w = 1$ and this value is not taken by $w = \sqrt{J(z^2)}$, we obtain the result that $w^* = p(z)$ maps $|z| < 1$ on a surface R_3 whose only boundary points are $w^* = 0, 1, \infty$ which are all outside R_3 , apart from $w^* = 0$ which is covered by one sheet of R_3 (for $z = 0$); moreover, R_3 contains no inner branch-points.

But this means that R_3 is identical with R_2 , as there is only one Riemann surface of this description. For let $w = J(z) = a_1z + a_2z^2 + \dots$ be the function mapping $|z| < 1$ on R_2 and $w = p(z) = b_1z + b_2z^2 + \dots$ ($a_1 > 0, b_1 > 0$) the function mapping $|z| < 1$ on R_3 . Clearly both functions $w(z) = J^{-1}[p(z)]$ and $w_1(z) = p^{-1}[J(z)]$ are regular for $|z| < 1$ and we have there $|w(z)| < 1$ and $|w_1(z)| < 1$, as no path within R_2 can lead out of R_3 and vice versa, both surfaces having only the isolated boundary points $w = 0, 1, \infty$ and no branch-points. The expansions of $w(z)$ and $w_1(z)$,

beginning with $w(z) = (b_1/a_1)z + \dots$ and $w_1(z) = (a_1/b_1)z + \dots$ respectively, we have $|b_1/a_1| \leq 1$ and $|a_1/b_1| \leq 1$, which, a_1 and b_1 being positive, implies $a_1 = b_1$ and consequently $w(z) \equiv z$ and $w_1(z) \equiv z$. Therefore, if we choose the positive branch of $\sqrt{J(z^2)}$, we are lead to the equation $J(z) = p(z)$, i. e.,

$$(1) \quad J(z) = 4\sqrt{J(z^2)}/[1 + \sqrt{J(z^2)}]^2.$$

As can be easily verified, this functional equation permits the successive computation of the coefficients a_n of the Taylor expansion $J(z) = \sum_{n=1}^{\infty} a_n z^n$. For the first coefficient a_1 we obtain $a_1 = 4\sqrt{a_1}$, i. e., $a_1 = 16$. This coefficient determines the constant in a classical theorem by Hurwitz to be mentioned later.

3. In order to find a convenient analytical expression for $J(z)$ and, at the same time, to establish the identity of $J(z)$ with the square of the modulus of the Jacobian elliptic functions, we proceed as follows:

Let $w = f(z; \rho)$ be the analytic function, real for real values of z , which maps the annulus $\rho < |z| < 1$ on the full circle $|w| < 1$ cut along the stretch $0 \leq w \leq \sqrt{\xi(\rho)}$. It can then be easily shown that the formalism connecting $\xi(\rho)$ with $\xi(\rho^2)$ is the same as that connecting $J(z)$ with $J(z^2)$. In fact, by inverting the annulus $\rho < |z| < 1$ along its outer boundary we find, by Schwarz' symmetry principle, that $w = f(z; \rho)$ maps the annulus $\rho < |z| < 1/\rho$ on the full w -plane cut along the stretch $0 \leq w \leq \sqrt{\xi(\rho)}$ and the ray $1/\sqrt{\xi(\rho)} \leq w \leq \infty$. Remembering that the function

$$w = 4/\sqrt{\xi(\rho)} \cdot x/(1+x)^2$$

maps the circle $|x| < 1$ on the full w -plane cut along the ray $1/\sqrt{\xi(\rho)} \leq w \leq \infty$, we have consequently

$$4f(\rho z; \rho^2)/\sqrt{\xi(\rho)}[1 + f(\rho z; \rho^2)]^2 = f(z; \rho).$$

For $z = \rho$, this reduces to

$$4f(\rho^2; \rho^2)/[1 + f(\rho^2; \rho^2)]^2 = \sqrt{\xi(\rho)} \cdot f(\rho; \rho).$$

But $f(z; \rho)$ being real for real values of z , we have necessarily

$$f(\rho; \rho) = \sqrt{\xi(\rho)}$$

and we thus obtain for $\xi(\rho)$ the functional equation

$$4\sqrt{\xi(\rho^2)}/[1+\sqrt{\xi(\rho^2)}]^2 = \xi(\rho),$$

i. e., equation (1).

Obviously, equation (1) is not only satisfied by the modular function $J(\rho)$ but also by all functions $J(\rho^\gamma)$, where γ is any complex number. In order to find out to which of these functions $\xi(\rho)$ corresponds, we need some more information about the analytical behavior of $\xi(\rho)$ (which so far is only defined for positive values of ρ) in the unit circle. To this end we shall construct an explicit expression for $f(z; \rho)$ which, for $z = \rho$, will furnish an expression for $\xi(\rho)$.

Since $w = f(z; \rho) = f(z)$ maps the circle $|z| = 1$ on the circle $|w| = 1$ and the circle $|z| = \rho$ on part of the real axis we have, by Schwarz' symmetry principle,

$$f(1/\bar{z}) = 1/\overline{f(z)}, \quad f(\rho^2/\bar{z}) = \overline{f(z)}.$$

Since $f(z)$ is real for real z , we have the additional relation $f(\bar{z}) = f(z)$ which, combined with the two previous ones, yields

$$(2a) \quad f(1/z) = 1/f(z)$$

and

$$(2b) \quad f(\rho^2/z) = f(z).$$

(2a) and (2b) entail the further relation

$$(2c) \quad f(\rho^4 z) = f(z);$$

indeed we have

$$f(\rho^4 z) = f(1/\rho^2 z) = 1/f(\rho^2 z) = 1/f(1/z) = f(z).$$

Accordingly, $g(x) = f(e^x)$ is an elliptic function possessing the two elementary periods $2\pi i$ and $4 \log \rho$.

$f(z)$ being real for real values of z , we have $f(-\rho) = 0$; moreover, $w = 0$ being the end of the cut $0 \leq w \leq \sqrt{\xi(\rho)}$, it follows that $z = -\rho$ is a double zero of $f(z)$. By (2b) and (2c), all further zeros of $f(z)$ are also double and coincide with the points $-\rho^{4n+1}$ ($n = 0, \pm 1, \pm 2, \dots$); obviously, there are no other zeros of $f(z)$ apart from these. By (2a), all poles of $f(z)$ are also double and coincide with the numbers $-\rho^{4n-1}$.

We now consider the infinite product

$$(3) \quad \phi(z) = z \cdot \frac{\prod_0^\infty (1 + \rho^{4n+1} z^{-1})^2 \prod_1^\infty (1 + \rho^{4n-1} z)^2}{\prod_0^\infty (1 + \rho^{4n-1} z^{-1})^2 \prod_1^\infty (1 + \rho^{4n+1} z)^2}, \quad (0 < \rho < 1),$$

which converges for all values of z differing from 0 and ∞ . By its construction, $\phi(z)$ has the same zeros and poles as $f(z)$. It is further easily verified that $\phi(z)$ satisfies the equation $\phi(\rho^4 z) = \phi(z)$, i. e., $\phi(e^x)$ is an elliptic function of x possessing the same zeros, poles and elementary periods as $f(e^x)$. By Liouville's theorem, $\phi(z)$ and $f(z)$ are therefore identical, apart from a multiplicative constant. In view of (2a), this constant can only take the values ± 1 . $f(z)$ being assumed positive for positive values of z , the negative sign is ruled out and we have

$$f(z) = \phi(z).$$

Remembering that $f(\rho) = \sqrt{\xi(\rho)}$, we obtain from (3):

$$\begin{aligned}\sqrt{\xi(\rho)} = f(\rho) = \phi(\rho) &= 4\rho \prod_{n=1}^{\infty} [(1 + \rho^{4n}) / (1 + \rho^{4n-2})]^4, \\ \xi(\rho) &= 16\rho^2 \prod_{n=1}^{\infty} [1 + \rho^{4n}) / (1 + \rho^{4n-2})]^8.\end{aligned}$$

This product converges absolutely for all values $|\rho| < 1$ and represents there an analytic function of ρ . As shown above, this function $\xi(\rho)$ satisfies for positive values of ρ the equation

$$\xi(\rho) = 4\sqrt{\xi(\rho^2)} / [1 + \sqrt{\xi(\rho^2)}]^2.$$

As a relation between analytic functions this equation must remain true for all values ρ for which the functions involved may be continued analytically, i. e., for all values $|\rho| < 1$.

As mentioned further above, the functional equation for $\xi(\rho)$ is also satisfied by all functions $\xi(\rho^\gamma)$. If we denote $\xi(\sqrt{\rho})$ by $\eta(\rho)$, we have for $\eta(\rho)$ the expression

$$\eta(\rho) = 16\rho \prod_{n=1}^{\infty} [(1 + \rho^{2n}) / (1 + \rho^{2n-1})]^8,$$

which shows that $\eta(\rho)$ is analytic for $|\rho| < 1$ and satisfies $\eta(0) = 0$, $\eta'(0) > 0$. $\eta(\rho)$ further satisfies equation (1). As there is only one such solution to equation (1), we must necessarily have

$$\eta(\rho) = J(\rho),$$

whence

$$(4) \quad J(z) = 16z \prod_{n=1}^{\infty} [(1 + z^{2n}) / (1 + z^{2n-1})]^8.$$

which is the well known formula for the square of the modulus of the Jacobian elliptic functions with $z = e^{\pi i \tau}$, where τ is the period ratio of these functions.

4. In addition to the functional equation (1), there are three more functional equations satisfied by $J(z)$, viz.,

$$(5) \quad J(-z) = J(z)/(J(z) - 1), \quad |z| < 1,$$

$$(6) \quad J(e^{-\pi x}) + J(e^{-\pi/x}) = 1, \quad R\{x\} > 0,$$

$$(7) \quad J(-e^{-\pi x}) \cdot J(-e^{-\pi/x}) = 1.$$

These equations are simple consequences of the fact that the set of points $(0, 1, \infty)$ is transformed into itself by the transformations $w^* = w/(w - 1)$, $w^* = 1 - w$ and $w^* = 1/w$ respectively.

As shown above, $w = J(z)$ maps $|z| < 1$ on a locally "schlicht" surface R_2 the only boundary points of which are $w = 0, 1, \infty$, all of them outside R_2 apart from $w = 0$ which is covered by one sheet of R_2 (for $z = 0$). This surface is obviously transformed into itself by the transformation $w^* = w/(w - 1)$ which leaves the origin fixed and interchanges the two points 1 and ∞ . The function mapping $|z| < 1$ on this surface is

$$w^* = J(z)/(J(z) - 1) = -J'(0)z + \dots$$

As seen earlier, a function $f(z)$ mapping $|z| < 1$ on R_2 is completely determined by the conditions $f(0) = 0$ and $f'(0) > 0$. As both $-J(z)/(J(z) - 1)$ and $-J(-z)$ satisfy these conditions, we must, therefore, have

$$J(-z) = J(z)/(J(z) - 1),$$

i. e., equation (5). This relation may also be derived directly from (1). From

$$J(z) = 4\sqrt{J(z^2)}/[1 + \sqrt{J(z^2)}]^2$$

follows

$$J(-z) = -4\sqrt{J(z^2)}/[1 - \sqrt{J(z^2)}]^2,$$

$\sqrt{J(z^2)}$ being an odd function of z . By combining these two formulae we obtain

$$\begin{aligned} J(z)/J(-z) &= -[(1 - \sqrt{J(z^2)})/(1 + \sqrt{J(z^2)})]^2 \\ &= 4\sqrt{J(z^2)}/[1 + \sqrt{J(z^2)}]^2 - 1 = J(z) - 1, \end{aligned}$$

i. e., (5).

In order to prove (6), we pass from R_2 to the surface R_1 described earlier, by the transformation

$$w = M(z) = J[e^{-\alpha([1+z]/[1-z])}], \quad \alpha > 0.$$

The surface R_1 on which $|z| < 1$ is mapped by $w = M(z)$ is locally "schlicht" and has no boundary points except $w = 0, 1, \infty$, all of which are outside R_1 . We now consider the function

$$w^* = \psi(z) = M(z) - \frac{1}{2} = J[e^{-\alpha([1+z]/[1-z])}] - \frac{1}{2},$$

where α ($\alpha > 0$) has been so chosen as to make $J(e^{-\alpha}) = \frac{1}{2}$. With this value of α we have $\psi(0) = 0$; besides, the surface on which $|z| < 1$ is mapped by $w^* = \psi(z)$ is symmetrical with regard to $w^* = 0$ (its boundaries being the points $w^* = \pm \frac{1}{2}, \infty$). We may, therefore, conclude that $\psi(z)$ is an odd function of z , i. e., $\psi(-z) = -\psi(z)$, or

$$(8) \quad J[e^{-\alpha([1+z]/[1-z])}] - \frac{1}{2} = -J[e^{-\alpha([1-z]/[1+z])}] + \frac{1}{2}.$$

With $x = (1+z)/(1-z)$ ($R\{x\} > 0$) this becomes

$$(8a) \quad J(e^{-\alpha x}) + J(e^{-\alpha/x}) = 1,$$

which is identical with (6) provided we have $J(e^{-\pi}) = \frac{1}{2}$ (as will be shown presently).

Equation (7) is obtained by combining (5) and (8a). Indeed, by (5):

$$J(-e^{-\alpha x}) \cdot J(-e^{-\alpha/x}) = J(e^{-\alpha x})J(e^{-\alpha/x})/[J(e^{-\alpha x}) - 1][J(e^{-\alpha/x}) - 1].$$

In view of (8a), the terms on the right-hand side cancel out and we obtain

$$(8b) \quad J(-e^{-\alpha x}) \cdot J(-e^{-\alpha/x}) = 1.$$

It now remains to show that $J(e^{-\pi}) = \frac{1}{2}$. To this end we note that, by (5), $J(e^{-\alpha}) = \frac{1}{2}$ entails $J(-e^{-\alpha}) = -1$. By virtue of (1) we have further

$$J(-ie^{-\alpha/2}) = 4\sqrt{J(-e^{-\alpha})}/[1 + \sqrt{J(-e^{-\alpha})}]^2 = 4i/(1+i)^2 = 2,$$

whence

$$J(e^{-\alpha})J(-ie^{-\alpha/2}) = 1.$$

Although formula (8b) does not generally mean that for any two values z_1 and z_2 for which $J(z_1) \cdot J(z_2) = 1$ there exists a value x such that $z_1 = -e^{-ax}$ and $z_2 = -e^{-a/x}$, we may show that in our case this is nevertheless true. In view of $J(-e^{-a}) = -1$, (8a) is fulfilled for $x = 1$. We now connect $w = -1$ with $w = \frac{1}{2}$ by a path t in the upper half-plane and let x vary continuously from 1 to the point defined by $-e^{-a-\pi i} = e^{-ax}$ so that $w = J(-e^{-ax})$ describes t . $w = 1/J(-e^{-ax}) = J(-e^{-a/x})$ will then describe a path t' in the lower half plane, likewise starting from $x = 1$ and ending at a point given by $-e^{-a/2+\pi i/2} = -e^{-a/x}$, since $-ie^{-a/2} = 2 = 1/J(-e^{-a-\pi i})$ and $J(z) = 2$ has no other solution except $z = -ie^{-a/2}$ inside the half circle $\text{Im}\{z\} \leq 0$, $|z| \leq e^{-a/2}$. (This will be shown in detail in the proof of Theorem V, Lemma II.) The equation

$$J(-e^{-a-\pi i}) \cdot J(-e^{-a/2+\pi i/2}) = 1$$

therefore entails the existence of a number x_0 such that

$$-\alpha - \pi i = -\alpha x_0$$

and

$$-\alpha/2 + \pi i/2 = -\alpha/x_0,$$

whence

$$(\alpha + \pi i)(\alpha - \pi i) = 2\alpha^2,$$

$$\alpha^2 + \pi^2 = 2\alpha^2,$$

$$\alpha^2 = \pi^2.$$

α being positive, we have therefore

$$\alpha = \pi.$$

5. Relation (6) also contains the fact that $J(e^{\pi i \tau})$, ($J_m\{\tau\} > 0$), is an automorphic function of τ . For if we write—as usual in the theory of elliptic functions— $k^2(\tau) = J(e^{\pi i \tau})$, (6) takes the form

$$(9) \quad k^2(\tau) + k^2[-(1/\tau)] = 1.$$

Because of $J(e^{\pi i \tau}) = J(e^{\pi i(\tau \pm 2)})$ we have further

$$(10) \quad k^2(\tau \pm 2) = k^2(\tau).$$

By combining (9) and (10) we obtain

$1 - k^2[-(1/\tau)] = k^2(\tau) = k^2(\tau \pm 2) = 1 - k^2[-(1/(\tau \pm 2))],$
i. e.,

$$k^2[-(1/\tau)] = k^2[-(1/(\tau \pm 2))].$$

On writing $-1/\tau$ instead of τ , this becomes

$$k^2(\tau) = k^2[\tau/(1 \pm 2\tau)].$$

Now it can be easily verified that by a suitable combination of the transformations $\tau \rightarrow \tau \pm 2$ and $\tau \rightarrow \tau/(1 \pm 2\tau)$ any transformation $\tau \rightarrow (\alpha\tau + \beta)/(\gamma\tau + \delta)$ can be obtained, where α and δ are two odd, and β and γ two even integers satisfying $\alpha\delta - \beta\gamma = 1$. For such integers $\alpha, \beta, \gamma, \delta$ we have, therefore,

$$k^2[(\alpha\tau + \beta)/(\gamma\tau + \delta)] = k^2(\tau).$$

PART II.

1. We now begin a discussion of a class of analytic functions, first considered by Hurwitz, which we shall denote by H . A function $w = f(z)$ $= a_1z + a_2z^2 + \dots$ is said to belong to H if $f(z)$ is regular in the unit circle and does not vanish there except for $z = 0$.

The elliptic modular function $J(z)$ clearly belongs to H ; moreover, any function $w = f(z)$ of H which, for $|z| < 1$, omits a finite value $w = d$, can be represented with the help of $J(z)$ and a bounded function $\omega(z)$. Indeed, if we consider the function

$$\omega(z) = J^{-1}[f(z)/d],$$

we see that, starting from $z = 0$, $\omega(z)$ may be uniformly continued within the whole circle $|z| < 1$, since the critical values of $J^{-1}[w]$ ($w = 0, 1, \infty$) are not taken by $w = f(z)/d$ for $0 < |z| < 1$. Besides, $|J^{-1}[w]| < 1$ for all values of w differing from $0, 1, \infty$. Hence,

$$(13) \quad f(z) = d \cdot J[\omega(z)],$$

where $\omega(z)$ is regular for $|z| < 1$ and

$$(13a) \quad |\omega(z)| \leq |z|,$$

i. e., $f(z)$ is subordinate to $d \cdot J(z)$.

From (13) and (13a), many properties of the functions $f(z)$ belonging to H may be inferred. The first result to be mentioned here is the following classical result due to Hurwitz:

THEOREM I. *A function $w = f(z) = z + a_2 z^2 + \dots$ belonging to H takes, for $|z| < 1$, every value inside the circle*

$$|w| = 1/16.$$

The constant $1/16$ is the best possible.

The proof of this theorem follows immediately from the representation (13), by which

$$1 = d \cdot \omega'(0) \cdot J'(0),$$

where d was a value not taken by $f(z)$ for $|z| < 1$. Remembering that $J'(0) = 16$, we have therefore $1 \leq 16 |d|$, i. e., $|d| \geq 1/16$.

Remark. It is worthy of note that in the representation $f(z) = d \cdot J[\omega(z)]$, $\omega(z)$ is not only bounded but satisfies also $\omega(z) \neq 0$ for $z \neq 0$; on the other hand, in the proof of Theorem I—as also in the proofs of other theorems to follow—only the fact that $\omega(z)$ is bounded is made use of. Consequently, these theorems also remain true for the more general class of functions $g(z) = d \cdot J[\omega_1(z)]$, where $\omega_1(z)$ is an otherwise unrestricted bounded function. This class of functions, say H_d , may also be characterized as follows: A function $w = g(z) = b_1 z + b_2 z^2 + \dots$ belongs to H_d if $g(z)$ is regular and $g(z) \neq d$ for $|z| < 1$ and if any zero $z = \alpha$ of $g(z)$ may be connected with $z = 0$ by a curve $z = z(t)$ such that the curve $w(t) = w[z(t)]$ does not surround the point $w = d$.

We therefore arrive at the following generalization of Theorem I:

THEOREM Ia. *Let $w = g(z) = z + b_2 z^2 + \dots$ be a function regular for $|z| < 1$; then any value $w = d$ not taken there by $w = g(z)$ and not surrounded by a curve $w(t) = g[z(t)]$, where the curve $z(t)$ connects $z = 0$ with another zero of $g(z)$, satisfies*

$$|d| \geq 1/16.$$

2. An immediate consequence of (13) is

$$|f(z)| \leq |d| \max_{|z|=\rho} |J(z)|.$$

Although the best possible, this inequality is not very convenient to handle. A more convenient bound for $|f(z)|$ —which, of course, is not “the best possible” any more—is given by

THEOREM II. *If $f(z)$ belongs to H and $f(z) \neq d$ for $|z| < 1$, then*

$$|f(z)| \leq (|d|/16)e^{-\pi^2/\log \rho}, \quad [|z| = \rho].$$

Proof. From the product expansion (4) which, by replacing z by $-z$, takes the form

$$(14) \quad -J(-z) = 16z \prod_{n=1}^{\infty} [(1 + z^{2n})/(1 - z^{2n-1})]^8,$$

we may conclude that all the coefficients of the expansion of $-J(-z)$ in powers of z are non-negative, this being true of each single factor $[(1 + z^{2n})/(1 - z^{2n-1})]$ making up this product. This entails the equality

$$\text{Max}_{|z|=\rho} |J(z)| = -J(-\rho)$$

and the inequality

$$-J(-\rho) \geq 16\rho.$$

In view of formula (7) which, replacing e^{-x} by ρ , may be written

$$J(-\rho)J(-e^{\pi^2/\log \rho}) = 1,$$

we, therefore, obtain

$$\begin{aligned} |f(z)| &\leq |d| \text{Max}_{|z|=\rho} |J(z)| = -|d| J(-\rho) \\ &= -(|d|)/[J(-e^{\pi^2/\log \rho})] \leq (|d|/16)e^{-\pi^2/\log \rho}. \end{aligned}$$

It should be noted that this inequality gives the correct order of growth of $\text{Max}_{|z|=\rho} |f(z)|$ for $\rho \rightarrow 1$, since $\lim_{\rho \rightarrow 0} -(J[-\rho]/16\rho) = 1$ entails

$$\lim_{\rho \rightarrow 1} -J(-\rho) \cdot 16e^{\pi^2/\log \rho} = 1.$$

3. By Theorem I, the modulus of the first coefficient of the Taylor expansion of $f(z)$ is smaller than or equal to the first coefficient of the expansion of $-|d|J(-z)$. We shall now show that the same inequality holds between all other corresponding coefficients of $f(z)$ and $-|d|J(-z)$.

THEOREM III. *Let $f(z) = \sum_{n=1}^{\infty} a_n z^n$ belong to H and let d be a value such that $f(z) \neq d$ for $|z| < 1$, then*

$$|a_n| \leq |d| A_n,$$

where A_n ($n = 1, 2, \dots$) are the coefficients of the expansion

$$-J(-z) = \sum_{n=1}^{\infty} A_n z^n.$$

Proof. As already seen, all coefficients A_n are non-negative. Furthermore, if we denote the first and second differences $A_n - A_{n-1}$ and $A_n - 2A_{n-1} + A_{n-2}$ by $D_n^{(1)}$ and $D_n^{(2)}$ respectively, we have, by (14),

$$-(1-z)J(-z) = \sum_{n=1}^{\infty} D_n^{(1)} z^n = 16z \frac{(1+z^2)^8}{(1-z)^7} \prod_{n=2}^{\infty} \left(\frac{1+z^{2n}}{1-z^{2n-1}} \right)^8$$

and

$$-(1-z)^2 J(-z) = \sum_{n=1}^{\infty} D_n^{(2)} z^n = 16z \frac{(1+z^2)^8}{(1-z)^7} \prod_{n=2}^{\infty} \left(\frac{1+z^{2n}}{1-z^{2n-1}} \right)^8$$

(with $D_1^{(1)} = A_1$, $D_1^{(2)} = A_1$, $D_2^{(2)} = A_2 - 2A_1$). Both products being composed of factors the expansion of which in powers of z have non-negative coefficients, all numbers $D_n^{(1)}$ and $D_n^{(2)}$ are non-negative. Now, by a well known theorem,¹ the coefficients a_n of a function $f(z) = \sum_{n=1}^{\infty} a_n z^n$ subordinate to another function $F(z) = \sum_{n=1}^{\infty} a'_n z^n$ with $a'_1 > 0$, $a'_2 - a'_1 \geq 0$, $a'_n - 2a'_{n-1} + a'_{n-2} \geq 0$ ($n \geq 3$), satisfy the inequality $|a_n| \leq a'_n$. This completes the proof.

Although Theorem III gives the exact bounds for the coefficients a_n , it is sometimes more desirable to have a convenient, though inexact, bound for the coefficients a_n , since the computation of the coefficients A_n becomes rather tedious for large n . Such a bound is given by

THEOREM IV. If $f(z) = \sum_{n=1}^{\infty} a_n z^n$ belongs to H and $f(z) \neq d$ for $|z| < 1$, we have

$$|a| \leq (|d|/16)e^{2\pi\sqrt{n}}.$$

This is an immediate consequence of Theorem II. Indeed:

$$|a_n| \leq |1/\rho^n| \max_{|z|=\rho} |f(z)| \leq (|d|/16)\rho^{-n} \cdot e^{-(\pi^2/\log \rho)}.$$

¹ See J. E. Littlewood, *Lectures on the Theory of Functions*, Oxford University Press, 1944, p. 169.

For $\rho = e^{-(\pi/\sqrt{n})}$ this becomes

$$|a_n| \leq (|d|/16)e^{2\pi\sqrt{n}}.$$

4. We shall now prove the following theorem:

THEOREM V. *If $f(z) = z + a_2z^2 + \dots$ belongs to H and $f(z) \neq \alpha$ for $|z| < 1$, then there exists a positive number $\rho = \rho(|\alpha|)$ such that $f(z)$ is univalent inside the circle $|z| = \rho$. For $|d| \geq e^\beta/16$, where β is the positive solution of*

$$\left(\frac{\sqrt{\beta+2} - \sqrt{\beta}}{\sqrt{\beta+2} + \sqrt{\beta}} \right) e^{-(\beta\sqrt{\beta}/[\beta+2])} = e^{-(\pi/2)},$$

we may take

$$\rho = 1 + \log 16 |d| - \sqrt{(1 + \log 16 |d|)^2 - 1},$$

and for $|\alpha| < e^\beta/16$, the positive solution of

$$\rho \cdot [16 |d|]^{(\rho-1)/(\rho+1)} = e^{-(\pi/2)}.$$

The value of ρ given for the case $|d| \geq e^\beta/16$ is the best possible, while in the case $|d| < e^\beta/16$ the value given tends to the best possible only for $|d| \rightarrow 1/16$.

For the proof of this theorem we require the following two lemmas:

LEMMA I. *Let $\omega(z) = \alpha z + c_2z^2 + \dots$ be regular and $|\omega(z)| \leq |z|$ for $|z| < 1$, and let $\omega(z)$ further satisfy $\omega(z) \neq 0$ for $0 < |z| < 1$; then $\omega(z)$ is univalent inside the circle $|z| = \rho$ with*

$$\rho = 1 + \log 1/|\alpha| - \sqrt{(1 + \log 1/|\alpha|)^2 - 1}.$$

This value of ρ is the best possible.

Proof. By hypothesis, $\omega(z)/z \neq 0$ and $|\omega(z)/z| \leq 1$ for $|z| < 1$. Hence, the function

$$g(z) = \log [z/\omega(z)] = \log 1/\alpha - (c_2/\alpha)z + \dots$$

is regular for $|z| < 1$ and we have there $R\{g(z)\} > 0$. By a classical result, this entails

$$(15) \quad R\{g(0)\}[(1-|z|)/(1+|z|)] \leq R\{g(z)\} \\ \leq R\{g(0)\}[(1+|z|)/(1-|z|)]$$

and

$$(16) \quad |g'(0)| \leq 2R\{g(0)\}.$$

In view of (15), we have

$$\log |z/\omega(z)| \geq \log 1/|\alpha| [(1-|z|)/(1+|z|)],$$

i. e.,

$$(17) \quad |\omega(z)| \leq |z| \cdot |\alpha|^{(1-|z|)/(1+|z|)}.$$

Since the linear substitution $z^* = (x-z)/(1-\bar{x}z)$ ($|x| < 1$) transforms the unit circle into itself, we have also $R\{g(z^*)\} > 0$. Now the first terms of the expansion of $h(z) = g(z^*)$ in powers of z are

$$h(z) = g[(x-z)/(1-\bar{x}z)] = g(x) - (1-|x|^2)g'(x)z + \cdots \\ = \log x/\omega(x) - \{(1-|x|^2)/x\} \{[1-(x\omega'(x))/\omega(x)]\}z + \cdots$$

The real part of $h(x)$ being positive for $|z| < 1$, we have, by (16),

$$|h'(0)| \leq 2R\{h(0)\},$$

whence

$$\{(1-|x|^2)/|x|\} |1-(x\omega'(x))/\omega(x)| \leq 2 \log |x/\omega(x)|,$$

i. e.,

$$(18) \quad |1-(x\omega'(x))/\omega(x)| \leq 2 \{|x|/(1-|x|^2)\} \log |x/\omega(x)|.$$

The necessary and sufficient condition for the circle $|z| < \rho$ to be mapped by $w = \omega(z)$ on a schlicht and star-shaped domain is $R\{z\omega'(z)/\omega(z)\} \geq 0$ for $|z| < \rho$. This condition is certainly fulfilled if $|1-(z\omega'(z)/\omega(z))| \leq 1$. By (18), this inequality holds for values z satisfying

$$\{2|z|/(1-|z|^2)\} \log |z/\omega(z)| \leq 1.$$

By (15), we have

$$\log |z/\omega(z)| \leq \log \{1/|\alpha|\} [(1+|z|)/(1-|z|)],$$

i. e.,

$$\{2|z|/(1-|z|^2)\} \log |z/\omega(z)| \leq \log 1/|\alpha| \cdot 2|z|/(1-|z|)^2.$$

$\omega(z)$ will therefore be univalent for $|z| < \rho$ with ρ satisfying

$$\log 1/|\alpha| \cdot 2\rho/(1-\rho)^2 = 1$$

i. e.,

$$\rho = 1 + \log 1/|\alpha| - \sqrt{(1 + \log 1/|\alpha|)^2 - 1}.$$

That this result is the best possible is shown by the function

$$\omega_0(z) = z \cdot e^{\log |a| [(1+z)/(1-z)]} = |a| z + \dots$$

for which $|\omega(z)| \leq |z|$ and $\omega(z) \neq 0$ for $0 < |z| < 1$, as required by the hypotheses of the lemma. As easily verified, the derivative of $\omega_0(z)$ vanishes at $z = \rho$ and $\omega_0(z)$ is therefore certainly not univalent in a circle $|z| < r$ with $r > \rho$.

LEMMA II. *The radius of univalence of the elliptic modular function $J(z)$ is $\rho = e^{-(\pi/2)}$.*

Proof. By (1), $J(z_1) = J(z_2)$ is equivalent to

$$4\sqrt{J(z_1^2)}/[1 + \sqrt{J(z_1^2)}]^2 = 4\sqrt{J(z_2^2)}/[1 + \sqrt{J(z_2^2)}]^2.$$

This equation has two solutions, viz., $J(z_1^2) = J(z_2^2)$ and $J(z_1^2) \cdot J(z_2^2) = 1$. If z_1 and z_2 are both assumed to lie on the circumference of the circle of univalence of $J(z)$, $J(z_1^2) = J(z_2^2)$ is clearly impossible and we must necessarily have $J(z_1^2) \cdot J(z_2^2) = 1$. As shown before, $|J(z)|$ attains its maximum on the circle $|z| = r$ for $z = -r$; we further had $J(e^{-\pi}) = \frac{1}{2}$ and $J(-z) = J(z)/(J(z) - 1)$, whence $J(-e^{-\pi}) = -1$. For values z satisfying $|z^2| < e^{-\pi}$ we therefore have $|J(z^2)| < 1$. Consequently, if $|z_1| < e^{-(\pi/2)}$ and $|z_2| < e^{-(\pi/2)}$, we certainly have $J(z_1^2) \cdot J(z_2^2) \neq 1$, which is equivalent to $J(z_1) \neq J(z_2)$. Therefore, $J(z)$ is univalent in the circle $|z| < e^{-(\pi/2)}$. On the other hand, in virtue of (1):

$$J(\pm ie^{-(\pi/2)}) = 4\sqrt{(-e^{-\pi})}/[1 + \sqrt{J(-e^{-\pi})}]^2 = \pm 4i/(1 \pm i)^2 = 2,$$

i. e., both $J(ie^{-(\pi/2)})$ and $J(-ie^{-(\pi/2)})$ have the value 2. Hence, $\rho = e^{-(\pi/2)}$ is the exact radius of univalence of $J(z)$.

In order now to prove Theorem V itself, we note that the univalence of $f(z) = d \cdot J[\omega(z)]$, ($|\omega(z)| \leq |z|$), can break down owing to two different reasons. These are: 1) $\omega(z)$ ceases to be univalent; 2) although $\omega(z_1) \neq \omega(z_2)$, we have $J[\omega(z_1)] = J[\omega(z_2)]$. Since, by Lemma II, the case $J[\omega(z_1)] = J[\omega(z_2)]$, $\omega(z_1) \neq \omega(z_2)$ cannot happen for $|\omega(z_1)| < e^{-(\pi/2)}$ and $|\omega(z_2)| < e^{-(\pi/2)}$ and, by Lemma I, $\omega(z) = z/16d + \dots$ is univalent for $|z| < \rho(|d|) = 1 + \log 16|d| - \sqrt{(1 + \log 16|d|)^2 - 1}$, $f(z)$ will also be univalent for $|z| < \rho(|d|)$. By (17), $|\omega(z)| \leq |z| [16|d|]^{1/2-1/|z|+1}$. Accordingly, the formula for the radius of univalence derived just now will

hold good as long as $|d|$ is larger than the positive solution $|d_0|$ of the equation obtained by eliminating ρ from the equations

$$\rho = 1 + \log 16 |d_0| - \sqrt{(1 + \log 16 |d_0|)^2 - 1}$$

and

$$\rho \cdot [16 |d|]^{\rho-1/\rho+1} = e^{-(\pi/2)},$$

i. e., $|d_0| = e^\beta/16$, where β is the positive solution of

$$\left(\frac{\sqrt{2+\beta} - \sqrt{\beta}}{\sqrt{2+\beta} + \sqrt{\beta}} \right) e^{-\beta\sqrt{\beta/(2+\beta)}} = e^{-(\pi/2)}.$$

If $|d|$ is smaller than $|d_0|$, the formula for the radius of univalence of $\omega(z)$ of course still holds, but since $|\omega(z)|$ may now be larger than $e^{-(\pi/2)}$, the multivalency of $J(z)$ may come into play. This can be avoided by choosing $|z|$ smaller than the positive solution ρ of

$$\rho [16 |d|]^{(\rho-1)/(\rho+1)} = e^{-(\pi/2)}$$

Since, for $|d| < |d_0|$, this value ρ is smaller than the radius of univalence of the functions $\omega(z)$, $f(z) = d \cdot J[\omega(z)]$ will therefore be univalent for $|z| < \rho$. This completes the proof.

THE HEBREW UNIVERSITY, JERUSALEM.

ON THE LAPLACE-FOURIER TRANSCENDENTS OCCURRING IN MATHEMATICAL PHYSICS.*

By AUREL WINTNER.

1. The success of Laplace's method of adjoints in case of a given differential equation (which is ordinary, linear, homogeneous and will be assumed to be of second order) always depends on the discovery of a "suitable" path of integration. The existence or non-existence of such paths is sensitive under arbitrarily small modifications of the coefficient functions. The situation is made particularly unstable by the implied demands of what are called the applications, which (for reasons of reality) prefer *rectilinear* paths and, at the same time, *positive* density functions (as to this terminology, cf. [6], pp. 226-240, 274-285). One could almost say that the relevant "classical transcendents of mathematical physics" have been added to the list of "known functions" because their differential equations present some of the few cases in which a contour and the method of contour deformations have, accidentally, succeeded in the above sense. Correspondingly, the only essential idea which, since the eighteenth century, has been added to the subject is Cauchy's theory of contour integrations, that is, his justification of Laplace's indiscriminate use of complex variables.

Thus it is natural to desire a replacement of the substantially finite set of classical examples by a theoretical approach which, upon an inspection of the differential equation, should be able to guarantee the existence of a solution representable as a definite integral of the type in question. Needless to say, such a plan cannot succeed if it involves, as usual, boundary conditions and contour deformations, the issue being precisely the existence or non-existence of an appropriate path of integration. Inasmuch as the desideratum is surely not novel (it could be as old as the method itself), it is somewhat unexpected that, as will be shown below, it can be realized, and in quite a "practicable" or "explicit" fashion, in *real* terms which are available since a long time (though not of course since Cauchy's time).

All that will be needed is a combination of two theories, both of which concern the real domain. The first of them is contained in A. Kneser's elementary considerations (1896) on the characteristics of differential equa-

* Received September 12, 1946.

tions of second order, made easily accessible by Borel's lectures on divergent series (1901). The second, available only since about a quarter of a century, is supplied by the existence theorem of Hausdorff-Bernstein on complete monotony, a result generally known. Correspondingly, what is somewhat unexpected is, not that the latter theorem, originally found in its own pure context, turns out to supply an answer to an old "applied" question, but the circumstance that this does not seem to have been observed before.

2. The above formulation of the desideratum is such as to make it clear that the criterion a priori cannot dispose of the necessity of explicit calculations, if explicit results are wanted. In this regard, the resulting situation can be described as follows: In the classical approach, there are two unknown elements; namely, a suitable integration path, which need not exist and must be guessed when it does, and a corresponding density function (the "lower function" of the transformation), which, if it exists, is supplied by a suitable solution of the adjoint differential equation. In contrast, the criterion a priori will assure the existence of a suitable integration path and of a corresponding density function. In addition, it will supply the determination of a suitable integration path, which will simply be the half-line $w = u$, ($0 < u < \infty$), in Laplace's case and the half-line $w = iv$, ($0 < v < \infty$), in Fourier's case. Moreover, the density function will be guaranteed to be real and non-negative. Since the possibility of its vanishing (though not, of course, of its identical vanishing) is included in its being non-negative, this will comprise the case of integration paths which are finite segments, or half-lines terminating at a point $z = c \neq 0$. But what will not (and, by the very nature of the desideratum, should not) be furnished is an explicit determination of that non-negative density function.

The calculation leading to such a determination must therefore remain dependent on a solution of the adjoint of the given differential equation or on equivalent means. Actually, the given differential equation being of second order, it will (at least in principle) be admissible to assume it in the normalized self-adjoint form,

$$(1) \quad x'' + f(t)x = 0.$$

The last parenthetical proviso is necessary, since any relevant criterion must surely be so sensitive as to fail to be covariant under transformations leading to the normal form (1); the more so as the coefficient function, say $g = g(t)$, which occurs in the general form, $(gx')'$, of a self-adjoint differential operation of second order is made to be $g \equiv 1$ (cf. Liouville's transformation). But the

normal form (1) suffices for the presentation of all the essential points in the procedure.

3. A systematic embedding of classical transcendents into the general theory will be developed elsewhere. The following remarks, illustrating the theory in the simplest case, that of Bessel's equation,

$$y'' + y'/r + (1 - \alpha^2/r^2)y = 0,$$

can therefore be somewhat high-handed.¹ In this case, one of the substitutions leading to the normal form (1) is simple indeed:

$$r = t, \quad y = t^{-\frac{1}{2}}x;$$

the coefficient of (1) then becomes

$$(1 \text{ bis}) \quad f(t) = 1 - \beta/t^2, \text{ where } \beta = (\alpha + \tfrac{1}{2})(\alpha - \tfrac{1}{2})$$

(for another, function-theoretically superior, substitution and the corresponding $f(t)$, cf. [4], p. 157). The index, α , is here arbitrary. However, the classical "real" representation of $J_\alpha(r)$ as ($r^{-\alpha}$ times) a Fourier transform is valid only when $\alpha > -\frac{1}{2}$. Instead of this, suppose that $\beta > 0$. Then, by the last formula line, $f(t)$ becomes the simplest of those polynomials in $1/t^2$ which have real coefficients with alternating signs (and with a positive zeroth coefficient).

It turns out that this property alone is sufficient for the existence of a solution of (1) which is a (generalized) Fourier transform of a non-negative density function. Moreover, nothing is changed if the polynomial is replaced by a transcendental entire function. In addition, even the requirement of an alternating coefficient sequence is unnecessarily strict. In fact, what it requires is that the coefficients should become positive after the substitution $t \rightarrow it$. But this is sufficient in order that $f(it)$ be completely monotone (in fact, the n -th derivatives of the functions $1/t^2, 1/t^4, 1/t^6, \dots$, where $0 < t < \infty$, are positive or negative according as n is even or odd). However, all that really matters is the complete monotony of the even function $f(it)$ on the half-line $t > 0$:

¹ What will be disregarded is the necessity of replacing the relevant solution, $y = J_\alpha(r)$, by the corresponding Fourier integral, which is $r^{-\alpha}$ times $J_\alpha(r)$; the additional transition to (1), which modifies the factor $r^{-\alpha}$ to $r^{-\alpha-\frac{1}{2}}$; finally, the determination of the "correct branch" when r is replaced by ri in this factor (the exponent, $-\alpha - \frac{1}{2}$, is an integer only in the elementary cases of J_α).

If $f(1/z)$ is an even entire function in z and if

$$(2) \quad (-1)^n d^n f(it)/dt^n \geq \text{when } 0 < t < \infty,$$

where $n = 0, 1, \dots$, (for instance, if

$$f(z) = \sum_{n=0}^{\infty} (-1)^n a_n / z^{2n} \text{ when } z \neq 0,$$

where $a_n \geq 0$), then the differential equation (1) possesses a Fourier solution $x(\neq 0)$, of the form

$$(3) \quad x(t) = \left(\int_0^{\infty} \right) e^{itr} d\phi(r) \text{ if } 0 < t < \infty,$$

where $\phi (\neq \text{const.})$ is a monotone, but not necessarily bounded, function on the half-line $0 \leq r < \infty$.

Since this half-line is closed, the possible unboundedness of ϕ takes place, of course, only when $r \rightarrow \infty$. The parenthetical integral sign in (3) refers to the Abelian value of the Fourier-Stieltjes transform. By this is meant that, if $t (> 0)$ is fixed, the integral

$$\int_0^{\infty} e^{-\epsilon r} d\phi_t(r), \text{ where } d\phi_t(r) = e^{itr} d\phi(r),$$

is convergent when $\epsilon > 0$ and tends, as $\epsilon \rightarrow 0$, to a finite limit, which is (3) (the existence of this limit is part of the assertion). According to the integral form of Abel's lemma on power series, (3) is identical with

$$\int_0^{\infty} e^{itr} d\phi(r), \quad (t > 0),$$

provided that the latter integral (which, since the total variation of $\phi(r)$ can be ∞ , must be meant as an improper Stieltjes, rather than as a Lebesgue-Stieltjes, integral) is convergent; in which case (3) is sure to exist. Incidentally, it remains undecided whether the replacement of the last formula line by (3) is actually necessary (for *some* f satisfying the assumptions of the theorem).

The assumptions required of f are satisfied if $f(t) \equiv 1$ (this is the limiting case, $\beta = 0$, of the normalized Bessel equation considered above). Then (1) becomes the differential equation $x'' + x = 0$, having the linearly

independent solutions $x = \cos t$, $x = \sin t$ which, since $f(t)$ is real, are equivalent to the single solution $x = e^{it}$. Inasmuch as the latter results by choosing $\phi(r)$ in (3) to be the step function $\frac{1}{2} \operatorname{sgn}(r-1)$, the appearance of a Stieltjes integral in this quite analytical context becomes understandable to some extent.

4. The function $\phi(r)$ must be absolutely continuous, in fact, piecewise analytic, if it can be obtained by the classical method of the adjoint differential equation. But, as exemplified by the function $J_0(t)$, a Fourier transform (3) can (even though, as shown by the example of $t^{-\frac{1}{2}}$ or $\exp(-t^2)$, need not) perform an infinity of "oscillations" or "waves," when t varies from 0 to ∞ . On the other hand, there are no *formal techniques* leading, by some method of comparing coefficients, to "*wave*" solutions of the type (3) (with $d\phi \geq 0$), when $f(t)$ in (1) is a given function. This is precisely the reason why all direct attempts trying to satisfy a (suitable, but still unspecified) differential equation (1) by anything like (3) could not succeed. However, the transformation $t \rightarrow it$, used above, replaces each of the partial vibrations, e^{irt} , of the bundle (3) by the corresponding damping, e^{-rt} , and the corresponding bundle cannot, of course, have "waves."

Actually, the solution curve which thus results upon the transformation $t \rightarrow it$ becomes so *smooth in the large* that a combination of the results of Kneser and of Hausdorff-Bernstein, referred to in the introduction, becomes applicable. Correspondingly, the criterion italicized above will not have to be considered directly, since it results as a straightforward corollary of the following theorem:

If a function $f(t)$, defined on the open half-line $t > 0$, has derivatives of arbitrarily high order satisfying

$$(4_n) \quad (-1)^n f^{(n)}(t) \leq 0, \quad 0 < t < \infty,$$

where $n = 0, 1, \dots$, then the differential equation (1) possesses a Laplacean solution $x(\neq 0)$, of the form

$$(5) \quad x(t) = \int_0^\infty e^{-tr} d\phi(r), \quad 0 < t < \infty,$$

where $\phi(\neq \text{const.})$ is a monotone, but not necessarily bounded, function on the closed half-line $r \geq 0$.

That this latter theorem implies the former as a corollary, follows from the circumstance that, if t in (5) is replaced by a complex variable, say z , then, since the convergence of the integral (5) on the half-line $t > 0$ is part of the latter theorem, the resulting integral converges to a regular function in the half-plane $\Re z > 0$. Thus the former theorem can be obtained by letting $\Re z \rightarrow 0$ and taking into account the definition of the parenthetical integral (3). In fact, every solution of a linear differential equation (1) (in which t is now complex) is regular at every point t at which the coefficient function, $f(t)$, is regular.

5. The second theorem is of a more symmetric nature, since it can simply be expressed as follows: *If $-f$ is completely monotone, then (1) has a completely monotone solution x ($\neq 0$); $0 < t < \infty$.* In fact, the existence theorem of Hausdorff-Bernstein states that (5), where $d\phi \geq 0$, is always implied by (and, of course, implies) the conditions

$$(6_n) \quad (-1)^n x^{(n)}(t) \geq 0, \quad 0 < t < \infty,$$

where $n = 0, 1, \dots$. The formulation in terms of (4_n) and (5) is preferable only because, f being the given function, condition (4_n) can easier be checked than the existence of an integral representation, whereas what is wanted of the unknown solution is precisely the existence of an integral representation.

If $x(t)$ is a completely monotone solution ($\neq 0$) of (1), another such solution is $Cx(t)$, where C is an arbitrary positive constant (this agrees with (5), where ϕ can be replaced by $C\phi$). But *only* this C remains undetermined: *Two completely monotone solutions of (1) are linearly dependent.* In fact, (5) implies that $x(t) = O(1)$ as $t \rightarrow \infty$. It does not imply that $x(t) = o(1)$, since $\phi(r)$ can have a jump at $r = 0$ (as illustrated by the simplest case, $f(t) \equiv 0$, which, according to (1), (5), belongs to $x(t) \equiv 0$, $\phi(r) = \text{sgn } r$). But it is clear that this effect of a possible jump of $\phi(r)$ at $r = 0$ is removed if $d\phi(r)$ is multiplied by r . Since the resulting integral is

$$\int_0^\infty e^{-tr} r d\phi(r) = -x'(t),$$

by (5), it follows that $x'(t) = o(1)$ as $t \rightarrow \infty$. Hence, if $x = x_1(t)$ and $x = x_2(t)$ are completely monotone, then since $x_1(t) = O(1)$ and $x_1'(t) = o(1)$ as $t \rightarrow \infty$, the Wronskian, $x_1 x_2' - x_2 x_1'$, is $o(1)$. However, the Wronskian of two solutions of (1) is always a constant and cannot, therefore, be $o(1)$ unless it vanishes identically. This proves the last italicized remark.

6. The theorem which remains to be proved is equivalent to the statement that, if f satisfies the infinity of conditions $(4_0), (4_1), \dots$, then (1) has a solution $x (\not\equiv 0)$ satisfying $(6_0), (6_1), \dots$. The proof will succeed only because the statement can be "finitized." In fact, it can be sharpened to the following assertion, in which m denotes a fixed non-negative integer:

(7_m) If a function $f(t)$, where $0 < t < \infty$, is of class $C^{(m)}$ and satisfies the $m + 1$ conditions $(4_0), \dots, (4_m)$, then the differential equation (1) has a solution $x (\not\equiv 0)$ satisfying the $m + 2$ conditions $(6_0), \dots, (6_{m+1})$.

This will be concluded, by complete induction, from the case $m = 0$, which is substantially equivalent to the following theorem:

(7*) If $f(t)$ is a real-valued, non-positive, continuous function on the open half-line $0 < t < \infty$, then the differential equation (1) has a solution $x(t)$ which is non-negative and non-increasing at every point of the half-line (and does not vanish everywhere).

Ad (7₀). The assumption of (7₀) is the same as that of (7*), namely, (4₀). The assertions of (7*) are (6₀) and (6₁), whereas (7₀) claims (6₂) also. However, since (1) means that

$$(8_0) \quad x'' = -fx,$$

(6₂) is implied by (6₀) and (7₀).

Ad (7₁). The assumptions of (7₁) are (4₁) (along with the existence of a continuous f') and the assumptions of (7₀). The new assertion is (6₃). However, from (8₀),

$$(8_1) \quad -x''' = fx' + f'x.$$

According to (4₀), (4₁) and (6₀), (6₁), both terms on the right of (8₁) are non-negative. Hence, (6₃) follows from (8₁).

Ad (7₂). The assumptions of (7₂) are (4₂) (along with the existence of a continuous f'') and the assumptions of (7₁). The new assertion is (6₄). However, from (8₁),

$$(8_2) \quad x'''' = -fx'' - 2f'x' - f''.$$

According to (4₀), (4₁), (4₂) and (6₀), (6₁), (6₂), all three terms on the right of (8₂) are non-negative. Hence, (6₄) follows from (8₂).

Since the binomial coefficients, introduced by m -fold differentiation of the product (8_0) , are positive, it is clear that the play of alternating signs is preserved in the induction $m \rightarrow m + 1$. Consequently, all that remains to be assured is (7^*) .

7. If the assumption, $f \leq 0$, of (7^*) is replaced by $f < 0$, then the assertion of (7^*) can be refined as follows:

(7 bis) If $g(t)$ is a positive, continuous function on the open half-line $0 < t < \infty$, then the differential equation

$$(9) \quad x'' = g(t)x$$

has a solution $x = x(t)$ which is positive (and so, by (9), convex) and decreasing for $0 < t < \infty$.

Remark. Neither

$$(10) \quad x(+0) \neq \infty$$

nor

$$(11) \quad x(\infty) \neq 0$$

is claimed in (7 bis).

Actually, (7^*) can be concluded from (7 bis), if use is made of Helly's selection theorem and of the locally continuous dependence of the general solution on $f = -g$.

Incidentally, (7 bis) instead of (7^*) , along with the induction $m \rightarrow m + 1$, is sufficient to assure the existence of a representation (5) in the case in which (4_n) is assumed for every n . In fact, $f(t)$ is regular at every t in this case, and so it is clear from the convexity assumption (4_2) that $f(t)$ cannot vanish at all unless it vanishes identically. But then the differential equation becomes $x'' = 0$ and possesses, therefore, the solution $x(t) \equiv 1$, which, of course, has the properties claimed.

Accordingly, only (7 bis) remains to be ascertained.

Let $(7')$ denote the sentence which becomes of the sentence (7 bis) if the open half-line, $0 < t < \infty$, is replaced by a closed half-line, say $1 \leq t < \infty$, in the assumption and in the assertion of (7 bis).

What corresponds to the negation of (10) (namely, the possibility $x(1+0) = \infty$) cannot occur in the case of a closed half-line. The latter can therefore be attacked by an initial point, $t = 1$. Thus (7 bis) is somewhat more elaborate than $(7')$. However, (7 bis) can be reduced to $(7')$. In fact, if an arc, say $t_1 \leq t \leq t_2$, of a solution path, $x = x(t)$, is in the upper

half-plane ($x > 0$), then it is convex ("subharmonic"); if it is in the lower half-plane ($x < 0$), it is concave. This is clear from (9), since $g > 0$. Hence, if (7') is granted, and if $x = x(t)$ denotes a solution supplied by (7') on the half-line $1 \leq t < \infty$, the extension of (7') to (7 bis) is furnished by that solution of (9) on the interval $0 < t \leq 1$ which is determined by the initial values $x(1)$, $x'(1)$.

This reduces everything to the truth of (7'). But (7') is known. In fact, (7') is contained in Kneser's geometrical considerations ([5], pp. 183-191), which deal with a non-linear generalization of (9), where $g > 0$. Correspondingly, Kneser's proof admits of simplifications in the linear case of (7') (in this respect, cf. pp. 42-46 of Borel's monograph [2]; actually, the proof can be simplified still further).

8. For the sake of completeness, a final comment must be made; one concerning the impossibility of (11), which is claimed by Kneser ([5], p. 191) but is not claimed above. This discrepancy can be cleared up as follows:

Let $x_1(t), x_2(t), \dots$ be a sequence of continuous (or, for that matter, regular analytic) functions on the half-line $1 \leq t < \infty$, and let t_1, t_2, \dots be a monotone sequence of t -values satisfying $1 < t_n \rightarrow \infty$ as $n \rightarrow \infty$. Suppose that $x_n(t)$ is negative and concave ("superharmonic") on the half-line $t_n < t < \infty$; that $x_n(t)$ is positive, convex and decreasing on the interval $1 \leq t < t_n$; finally, that all curves $x = x_n(t)$ meet when $t = 1$ (and, for the sake of simplicity, that, if n and m are distinct, $x_n(t) = x_m(t)$ holds only when $t = 1$). Then it is clear, for instance from Helly's theorem, that the sequence $x_1(t), x_2(t), \dots$ contains a subsequence which, uniformly on every bounded t -interval, tends to a limit function, say $x(t)$; and that $x(t)$ is positive, convex and decreasing on the whole half-line. In particular, $x(t)$ tends to a finite, non-negative limit as $t \rightarrow \infty$.

What Kneser seems to consider to be evident ([5], p. 182) is that, since $t_n \rightarrow \infty$ as $n \rightarrow \infty$ and $x_n(t_n) = 0$, this non-negative limit, $x(\infty)$, cannot be positive. Actually, it is easy to draw a sequence of curves which satisfy all the above conditions but lead to a positive $x(\infty)$. In fact, the convergence, uniform on every bounded t -interval of the half-line, need not be uniform on the whole of the half-line.

This explains the discrepancy. It was mentioned because it is important in connection with (3), (5). In fact, the inclusion of discontinuous functions $\phi(r)$ was seen to be an essential point in the above theorems on integral representations. But (11) is precisely the condition for a solution (5) in which $\phi(r)$ has no jump at $r = 0$.

Since the unknown of the problem is the function $\phi(r)$, it is worth while to mention that a jump $\phi(+0) \neq \phi(0)$ in (5) can be excluded by a criterion involving only the coefficient function of (1). In fact, such a criterion is contained in the following remark (which, instead of (4_n), assumes only that $f(t) \leq 0$):

If $f(t)$ and $x(t)$ satisfy the assumptions of (γ^), then (11) cannot be true unless the integral*

$$(12) \quad \int_0^\infty f(t) dt \text{ is convergent.}$$

For, if $f(t)$ and $x(t)$ satisfy the assumptions of (γ^*), then, since $x(t)$ is convex by virtue of (1), the derivative $x'(t)$ must tend to 0 as $t \rightarrow \infty$. Since (1) means that

$$(13) \quad x'(t) + \int_0^t f(u)x(u) du$$

is independent of t , it follows that

$$(14) \quad \int_0^\infty f(t)x(t) dt \text{ is convergent,}$$

whether (12) be satisfied or not. Hence, it is clear from $f(t) \leq 0$ that (11) is impossible if (12) is assumed (the existence of $x(\infty)$ being implied by (γ^*) alone).

Appendix.

Under the assumptions which had to be made for the existence of Fourier and Laplace solutions, (3) and (5), the coefficient function, $f(t)$, of (1) must of course be non-negative and non-positive, respectively. Correspondingly, the simplest illustrations of the "elliptic" case and of the "hyperbolic" case are $f(t) = a^2$ with $x(t) = \cos at$ and $f(t) = -a^2$ with $x(t) = e^{-at}$, respectively, where a is a positive constant; cf. (3), (5) and (1). In the case of e^{-at} , the possibility (11) does not take place. In the case of $\cos at$, the (real) solutions $x(t)$ are oscillatory, that is to say such as to possess an infinity of zeros (of first order). The limiting "parabolic" case, $f(t) \equiv 0$, of (1) satisfies the assumptions of both (5) and (3). But since all solutions of $x'' = 0$ are linear, each of its solutions $x(t) \not\equiv 0$ presents the exceptional possibility (11) and none of them is oscillatory. Hence, if (12) is interpreted

as an assumption the negation of which carries (1) "too close" to $x'' = 0$ in the hyperbolic case, $f(t) \leq 0$, of (7*), it is reasonable to expect that the same negation will prevent the existence of non-oscillatory solutions in the elliptic case, $f(t) \geq 0$.

This heuristic dualization suggests a counterpart of the last italicized remark. It is easy to see that the resulting dual happens to be true:

If $f(t)$, where $0 < t < \infty$, is a real-valued, non-negative, continuous function, then, unless (12) is satisfied, every real-valued solution $x(t) \not\equiv 0$ of (1) is oscillatory.

This remark, though quite on the surface (and possibly well-known), is not clear from a direct "comparison" (Sturm) of (1) with $x'' + a^2x = 0$. In fact, the negation of (12), that is, the assumption

$$(15) \quad \int_0^{\infty} f(t) dt = \infty, \text{ where } f(t) \geq 0,$$

is compatible with $\lim f(t) = 0$ (as is (12) with $\limsup |f(t)| = \infty$), as $t \rightarrow \infty$.

Since an interval containing two zeros of any (real-valued) solution $x(t) \not\equiv 0$ of (1) must contain a zero of any other such solution of (1) (Sturm), and since $x(t)$ can be replaced by $-x(t)$ in (1), it is sufficient to verify the following assertion: If (1) has a solution which is positive from a certain $t = t_0$ onward, then $f(t)$ cannot satisfy (15).

To this end, let the integral condition be disregarded. The remaining assumptions (namely, $f(t) \geq 0$ and $x(t) > 0$, where $t > t_0$) imply, by (1), that $x''(t) \leq 0$. In other words, the curve $x = x(t)$ is concave from below (from $t = t_0$ onward). Since it stays in the upper half-plane, it follows that $x(t)$ is non-negative and non-decreasing, whereas $x'(t)$ is non-negative and non-increasing. Hence, there exist a positive limit $x(\infty) \leq \infty$ and a non-negative limit $x'(\infty) < \infty$. But (13) is independent of t by virtue of (1), and so the existence of a finite limit $x'(\infty)$ implies (14). Finally, (14) and the existence of a (finite or infinite) positive limit $x(\infty)$ necessitate (12), since $f(t) \geq 0$.

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES

-
- [1] S. Bernstein, "Sur les fonctions absolument monotones," *Acta Mathematica*, vol. 52 (1928-29), pp. 1-66.
 - [2] E. Borel, *Leçons sur les séries divergentes*, Paris, 1901.
 - [3] F. Hausdorff, "Summationsmethoden und Momentfolgen, II," *Mathematische Zeitschrift*, vol. 9 (1921), pp. 280-299.
 - [4] F. Klein, *Vorlesungen über die hypergeometrische Funktion* (ed. O. Haupt), Berlin, 1933.
 - [5] A. Kneser, "Untersuchung und asymptotische Darstellung der Integrale gewisser Differentialgleichungen bei grossen reellen Werthen des Arguments, I," *Journal für die reine und angewandte Mathematik*, vol. 116 (1896), pp. 178-212.
 - [6] L. Schlesinger, *Einführung in die Theorie der gewöhnlichen Differentialgleichungen auf funktionentheoretischer Grundlage*, Berlin and Leipzig, 1922 (3rd ed.).

ON THE TAUBERIAN NATURE OF IKEHARA'S THEOREM.*

By AUREL WINTNER.

In case of ordinary Dirichlet series, Ikehara's theorem can be formulated as follows: If the series

$$(1) \quad f(s) = \sum_{n=1}^{\infty} a_n/n^s$$

is absolutely convergent in the half-plane $\sigma > 1$ and represents there a function which goes over into a continuous boundary function on the line $\sigma = 1$, then the Tauberian condition

$$(2) \quad a_n = O_L(1)$$

is sufficient to ensure that

$$(3) \quad \sum_{n=1}^x a_n = o(x).$$

This is not the usual formulation (cf. [6], pp. 127-130) but is readily seen to be equivalent to it. In fact, if a constant, c , is added to every a_n , then (2) remains unaltered, (1) goes over into $f(s) + c\zeta(s)$ and acquires therefore a "pole", of "residue" c , at $s = 1$, and, correspondingly, the $o(x)$ in (3) becomes $cx + o(x)$.

The proof of Ikehara's theorem depends on the same technique as Wiener's result on Lambert summability. The content of the latter result is (cf. [6], pp. 119-124) that the (L) -summability of a series $c_1 + c_2 + \dots$ and the Tauberian condition

$$(2 \text{ bis}) \quad c_n/n = O_L(1)$$

imply (A) -summability. The analogy is the more relevant as either of the theorems suffices for the transition from the non-vanishing of $\zeta(1 + it)$ to the prime number theorem. It is true that, in the Lambertian approach, recourse must be had to the Hardy-Littlewood lemma, according to which (A) -summability and (2 bis) together imply $(C, 1)$ -summability (in fact, convergence as well; but this is not needed in the deduction of the prime number theorem), but Karamata [4] has discovered that this additional step

* Received September 12, 1946.

is simple indeed. For the general methodical situation, cf. [1], pp. 34-35, 46-47 and [3], pp. 38-40.

Actually, Wiener's Tauberian result on (L) -summability, when formulated in the above form, somewhat disguises the true situation. In fact, his result is just a corollary of an earlier theorem of Hardy and Littlewood [2], according to which (L) -summability *always* implies (A) -summability (whether a Tauberian restriction, such as (2 bis), be satisfied or not). But this theorem, though deeper than the prime number theorem itself, is not at all Tauberian in nature.

It is therefore natural to ask whether the situation is similar in the parallel case of Ikehara's approach to the prime number theorem; in other words, whether his Tauberian restriction, (2), is actually superfluous (at any rate, by virtue of the prime number theorem or, as the unconditional implication, $(L) \rightarrow (A)$, of Hardy and Littlewood, by virtue of a somewhat deeper theorem). The purpose of this note is *to answer this question in the negative*.

The situation would be simple indeed if, in the above formulation of Ikehara's theorem (which, as it stands above, is tautologous), it were not stipulated that the Dirichlet series be *absolutely* convergent in the half-plane $\sigma > 1$. For, if

$$a_n = (-1)^n,$$

the Dirichlet series (1) is convergent in the half-plane $\sigma > 1$ and attains continuous boundary values on the line $\sigma = 1$; in fact, the function (1) becomes the entire function $(2^{2-s} - 1)\zeta(s-1)$. Nevertheless, since

$$\liminf_{x \rightarrow \infty} \sum_{n=1}^x (-1)^n/x = -\frac{1}{2}, \quad \limsup_{x \rightarrow \infty} \sum_{n=1}^x (-1)^n/x = \frac{1}{2},$$

(3) is not true.

The construction of a Dirichlet series (1) having 1 as its abscissa of *absolute* convergence will be based on the general identity on which Riemann's proofs of the functional equation of $\zeta(s)$ depend. This identity, first justified by Perron [5], states that, if (1) is convergent in the half-plane $\sigma > 1$, then

$$(4) \quad f(s) = F(s)/\Gamma(s),$$

where

$$(5) \quad F(s) = \int_0^\infty x^{s-1} \sum_{n=1}^\infty a_n e^{-nx} dx, \quad \sigma > 1.$$

Needless to say, the convergence of (1) for $\sigma > 1$ implies that the power series multiplying x^{s-1} in (5) is convergent for $e^{-x} < 1$, i. e., for every x in the *interior* of the integration domain (the integral (5) must therefore be interpreted as improper at both ends, $x = \infty$ and $x = 0$).

Suppose that

$$(6) \quad \int_0^{1-0} \left| \sum_{n=1}^{\infty} a_n r^n \right| dr < \infty.$$

Then the integral (5) is absolutely, hence uniformly, convergent in the closed strip $1 \leq \sigma \leq 2$. In fact, this is clear for the contribution,

$$\int_1^{\infty} x^{s-1} \sum_{n=1}^{\infty} a_n e^{-nx} dx.$$

of the integration range $1 \leq x < \infty$. On the other hand, the contribution of the remaining range, $0 \leq x \leq 1$, to (5) is majorized, uniformly in the closed half-plane $\sigma \geq 1$, by

$$\int_0^{\infty} \left| \sum_{n=1}^{\infty} a_n e^{-nx} \right| dx = \int_0^{1-0} \left| \sum_{n=1}^{\infty} a_n r^n \right| dr/r, \quad (r = e^{-x}).$$

But the last integral differs from the integral (6) only in the factor $1/r$, which does not influence the convergence assumed for (6), since $1/r$ is bounded near $r = 1$, and is absorbed by r^n at $r = 0$, the summation index $n = 0$ being excluded.

Consequently, the function (5) is uniformly continuous (and, incidentally, bounded) in the open strip $1 < \sigma < 2$. Since $1/\Gamma(s)$ is not bounded in this strip, it does not follow that the function (4) is uniformly continuous there. But, since $\Gamma(1 + it) \neq 0$, it does follow that the function (4) represents in the half-plane $\sigma > 1$ a regular function which attains continuous boundary values on the line $\sigma = 1$.

Accordingly, it is sufficient to show that the following three conditions are compatible for a sequence a_1, a_2, \dots :

- (i) The Dirichlet series (1) is absolutely convergent when $\sigma > 1$.
- (ii) Condition (6) is satisfied.
- (iii) The estimate (3) fails to hold.

To this end, first choose an increasing sequence of positive integers k_1, k_2, \dots in such a way that, on the one hand,

$$(7) \quad \sum_{m=1}^{\infty} k_m^{-\epsilon} < \infty \text{ for every } \epsilon > 0$$

and, on the other hand,

$$(8) \quad \sum_{m=1}^{\infty} k_m r^{k_m} = O(1-r)^{-3/2} \text{ as } r \rightarrow 1-0.$$

Such sequences exist (in fact, both conditions are satisfied by $k_m = m!^2$, the second, (8), being satisfied even by $k_m = m^2$).

It can also be assumed that the sequence k_1, k_2, \dots does not contain consecutive integers. Then a unique sequence a_1, a_2, \dots is defined by the following assignment: a_n is n , $-n+1$ or 0 according as n is a k_m , a k_m+1 or neither a k_m nor a k_m+1 .

It is readily seen from this definition of a_n and from the first of the conditions, (7), imposed on the sequence k_1, k_2, \dots , that (i) is satisfied. In order to ascertain (ii), it is sufficient to show that

$$(9) \quad \sum_{n=1}^{\infty} a_n r^n = O(1-r)^{-1/2} \text{ as } r \rightarrow 1-0.$$

But the definition of a_n means that the power series on the left of (9) is identical with

$$\sum_{m=1}^{\infty} k_m r^{k_m} - \sum_{m=1}^{\infty} k_m r^{k_m+1};$$

so that, since this difference can be contracted into

$$(1-r) \sum_{m=1}^{\infty} k_m r^{k_m},$$

(9) follows from (8).

Accordingly, only (iii) remains to be verified. However, by the definition of a_n ,

$$(10) \quad \sum_{n=1}^x a_n = \sum_{k_m \leq x} k_m - \sum_{k_m \leq x-1} k_m.$$

Clearly, the difference on the right of (10) vanishes except at x -values

corresponding to its points of discontinuity, and the jumps amount to just the respective values of x itself. Accordingly,

$$\liminf_{x \rightarrow \infty} \sum_{n=1}^x a_n/x = 0 \text{ but } \limsup_{x \rightarrow \infty} \sum_{n=1}^x a_n/x = 1,$$

and so (3) fails to be true.

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES

-
- [1] G. H. Hardy, *Ramanujan*, Cambridge, 1940.
 - [2] G. H. Hardy and J. E. Littlewood, "On a Tauberian theorem for Lambert's series, and some fundamental theorems in the analytic theory of numbers," *Proceedings of the London Mathematical Society*, ser. 2, vol. 19 (1921), pp. 21-29.
 - [3] A. E. Ingham, *The Distribution of Prime Numbers*, Cambridge Tracts no. 30, 1932.
 - [4] J. Karamata, "Über die Hardy-Littlewoodsche Umkehrung des Abelschen Stetigkeitssatzes," *Mathematische Zeitschrift*, vol. 32 (1930), pp. 319-320.
 - [5] O. Perron, "Zur Theorie der Dirichletschen Reihen," *Journal für die reine und angewandte Mathematik*, vol. 134 (1908), pp. 95-143.
 - [6] N. Wiener, *The Fourier Integral and Certain of its Applications*, Cambridge, 1933.

UNION-PRESERVING TRANSFORMATIONS OF HIGHER ORDER SURFACE-ELEMENTS.*

By JOHN DE CICCIO.

1. Curve-element transformations of the union-preserving type. In 1943, Kasner and De Ciccio began the study of transformations in the plane and in space from differential elements of order n into those of first order.¹ We have termed such a correspondence *union-preserving* if it converts every union into a union. This class of union-preserving transformations includes the Lie theory of contact transformations as a special case.² In the previous work, our attention was devoted to the study of correspondences from curve-elements of order n into lineal-elements.³ In the present article, we shall develop the theory of correspondences from surface-elements of order n into planar-elements.

For the purpose of contrasting our new theory with the previous study,⁴ we shall state some of the results concerning union-preserving transformations in space from curve-elements of order n : $(x, y, z, y', z', \dots, y^{(n)}, z^{(n)})$, where the accents denote total differentiation with respect to x , into lineal-elements: (X, Y, Z, Y', Z') . The general union-preserving transformations are determined by a single directrix equation of the form

$$(C) \quad \Omega(X, Y, Z, x, y, z, y', z', \dots, y^{(n-2)}, z^{(n-2)}) = 0,$$

involving derivatives of order $(n-2)$, at most. If $n=2$, this has the same form as the directrix equation of Lie. Although the directrix equations are the same for this case $n=2$, the induced union-preserving transformation of second order is different in general from the induced contact transformation of Lie.

The special union-preserving transformations are obtained by extending

* Received August 15, 1946. Presented to the American Mathematical Society, 1946.

¹ Kasner and De Ciccio, "Union-preserving transformations of differential elements," *Proceedings of the National Academy of Sciences*, vol. 29 (1943), pp. 271-275.

² Lie-Scheffers, *Berührungstransformationen*, Leipzig, 1896.

³ Kasner and De Ciccio, "A generalized theory of contact transformations," *Rivista de Matematicas de la Universidad Nacional de Tucuman*, Argentina, vol. 4 (1944), pp. 81-90.

⁴ Kasner and De Ciccio, "Union-preserving transformations of space," *Bulletin of the American Mathematical Society*, vol. 50 (1944), pp. 98-107. "Comparison of union-preserving and contact transformations," *Proceedings of the National Academy of Sciences*, vol. 32 (1946), pp. 152-156.

to lineal-elements transformations which carry curve-elements of order $(n-1)$ into points. Thus the class of union-preserving transformations from curve-elements of order n into lineal-elements consists of the general type defined by a single directrix equation of the form (C) or of the special type described above.

Any transformation not of the union-preserving type converts either precisely $\infty^{2(n+1)}$ or ∞^∞ unions into unions.

The only available union-preserving transformations in the whole domain of curve-elements are firstly, the group of point transformations, and secondly, the set of union-preserving transformations from curve-elements of order n into lineal-elements; together with the extensions of these two types. This is an extension of Lie's theorem which states that the group of contact transformations of lineal-elements constitutes the extensions of the correspondences of the group of arbitrary point transformations.

2. Union-preserving transformations of higher order surface-elements.

In this section, we shall state some of the results of the present article. It will be found that our new transformation theory is more analogous to the Lie theory than that described in the preceding section.

We shall study transformations in space from differential surface-elements of order n : $(x, y, z, p_{10}, p_{01}, \dots, p_{j,m-j}, \dots, p_{0n})$, where $p_{j,m-j} = \partial^m z / \partial x^j \partial y^{m-j}$ for $j=0, 1, 2, \dots, m$, and $m=1, 2, \dots, n$, into planar-elements (X, Y, Z, P, Q) where $P = \partial Z / \partial X$ and $Q = \partial Z / \partial Y$. Thus our transformations convert the $\infty^{(n^2+3n+6)/2}$ surface-elements of order n into the ∞^5 planar-elements. Our union-preserving transformations may be defined by a single directrix equation

$$(S) \quad \Omega(X, Y, Z, x, y, z, p_{10}, p_{01}, \dots, p_{j,m-j}, \dots, p_{n-1,0}, \dots, p_{0,n-1}) = 0,$$

involving partial derivatives of order $(n-1)$ at most; or by a pair or triplet of such directrix equations. For $n=1$ this result coincides with the theorem of Lie concerning contact transformations of surface-elements of first order.

Thus our union-preserving transformations are defined by one arbitrary function of $(n^2+n+8)/2$ independent variables, or by two arbitrary functions of $(n^2+n+6)/2$ independent variables, or by three arbitrary functions of $(n^2+n+4)/2$ independent variables. Therefore in the notation introduced by Kasner, the contents of the three different classes of union-preserving transformations are $\infty^{1f[(n^2+n+8)/2]}$, $\infty^{2f[(n^2+n+6)/2]}$, and $\infty^{3f[(n^2+n+4)/2]}$, respectively.⁵

⁵ Kasner, "A notation for infinite manifolds," *American Mathematical Monthly*, vol. 49 (1942), pp. 243-244.

Any transformation not of the union-preserving type converts at most $\infty^{(n^2+5n+2)/2}$ or ∞^∞ unions into unions. For $n=1$, this coincides with a theorem of Kasner concerning transformations of surface-elements of first order.

The only available union-preserving transformations in the whole domain of surface-elements are firstly, Lie's group of contact transformations of planar-elements, and secondly, our set of union-preserving transformations from surface-elements of order n into planar-elements, together with the extensions of these two types.

3. The conditions for union-preserving transformations. Any transformation T from surface-elements of order n into planar-elements is defined by equations of the forms

$$(1) \quad \begin{aligned} X &= X(x, y, z; p_{j,m-j}), & Y &= Y(x, y, z; p_{j,m-j}), & Z &= Z(x, y, z; p_{j,m-j}), \\ P &= P(x, y, z; p_{j,m-j}), & Q &= Q(x, y, z; p_{j,m-j}), \end{aligned}$$

where

$$(2) \quad p_{j,m-j} = \partial^m z / \partial x^j \partial y^{m-j} \text{ for } j = 0, 1, 2, \dots, m, \text{ and } m = 1, 2, \dots, n.$$

Of course, $P = \partial Z / \partial X$ and $Q = \partial Z / \partial Y$.

Thus any transformation T converts the totality of $\infty^{(n^2+3n+6)/2}$ surface-elements of order n of a certain region of the (x, y, z) -space into the ∞^5 planar-elements of a certain region of the (X, Y, Z) -space provided that the rank of the jacobian-matrix

$$(3) \quad \begin{bmatrix} X_x & Y_x & Z_x & P_x & Q_x \\ X_y & Y_y & Z_y & P_y & Q_y \\ X_z & Y_z & Z_z & P_z & Q_z \\ X_{p_{j,m-j}} & Y_{p_{j,m-j}} & Z_{p_{j,m-j}} & P_{p_{j,m-j}} & Q_{p_{j,m-j}} \end{bmatrix}$$

which consists of five columns and $(n^2 + 3n + 6)/2$ rows, is five throughout a certain region of the (x, y, z) -space.

A *double-series* consists of ∞^2 surface-elements of order n and may be defined by the parametric equations: $x = x(r, s)$, $y = y(r, s)$, $z = z(r, s)$; $p_{j,m-j} = p_{j,m-j}(r, s)$, where r and s are independent parameters, such that at least two of the functions are functionally independent. If all the functions in this system of equations are functionally related, the surface-elements of order n reduce to ∞^1 in number, and the resulting geometric configuration is called a *simple series*.

A double-series forms a *union* if and only if

$$(4) \quad dz = p_{10}dx + p_{01}dy, \quad dp_{j,m-j} = p_{j+1,m-j}dx + p_{j,m-j+1}dy,$$

for $j = 0, 1, 2, \dots, m$, and $m = 1, 2, \dots, n-1$. A simple series is termed a *strip* if the system of equations defining it satisfy the above system (4) identically.

A special type of union is the *conical-union*. This consists of ∞^2 surface-elements of order n which have in common a fixed surface-element of order $(n-1)$. The equations of any conical union are: $x = x_0, y = y_0, z = z_0, p_{10} = (p_{10})_0, p_{01} = (p_{01})_0, \dots, p_{n-1,0} = (p_{n-1,0})_0, \dots, p_{0,n-1} = (p_{0,n-1})_0, p_{n,0} = p_{n,0}(r, s), \dots, p_{0,n} = p_{0,n}(r, s)$, where r and s are independent parameters. Through a given surface-element of order $(n-1)$, there pass $\infty^{(n-1)f(2)}$ conical-unions. Thus in totality, there are $\infty^{(n^2-n+4)/2+(n-1)f(2)}$ conical-unions.

Before continuing with our work, it is found convenient to introduce the two linear operators

$$(5) \quad \begin{aligned} A_k &= \partial/\partial x + p_{10}(\partial/\partial z) + \sum_{m=1}^k \sum_{j=0}^m p_{j+1,m-j}(\partial/\partial p_{j,m-j}), \\ B_k &= \partial/\partial y + p_{01}(\partial/\partial z) + \sum_{m=1}^k \sum_{j=0}^m p_{j,m-j+1}(\partial/\partial p_{j,m-j}). \end{aligned}$$

It follows that if λ is any function of $(x, y, z; p_{j,m-j})$ involving partial derivatives of order k at most, then $A_k(\lambda)$ (or $B_k(\lambda)$) represents the total partial derivative of λ with respect to x (or y) where z is considered to depend on both x and y . It is found that both $A_k(\lambda)$ and $B_k(\lambda)$ involve partial derivatives of order $(k+1)$, at most.

We seek all the unions that become unions under the transformation T as defined by the system of equations (1). By (1), (3), and (5), we find that the required unions satisfy the system of partial differential equations

$$(6) \quad A_n(Z) = PA_n(X) + QA_n(Y), \quad B_n(Z) = PB_n(X) + QB_n(Y),$$

of order $(n+1)$. In general, there are at most $\infty^{(n^2+5n+2)/2}$ solutions except when the equations are in involution in which case the solutions involve arbitrary functions.⁶

THEOREM 1. *All the union-preserving transformations T from surface-elements of order n into planar-elements satisfy the $(n+3)$ conditions*

$$(7) \quad \begin{aligned} Z_{p_{j,n-j}} &= PX_{p_{j,n-j}} + QY_{p_{j,n-j}} \text{ for } j = 0, 1, 2, \dots, n, \\ A_{n-1}(Z) &= PA_{n-1}(X) + QA_{n-1}(Y), \\ B_{n-1}(Z) &= PB_{n-1}(X) + QB_{n-1}(Y). \end{aligned}$$

⁶ Kasner, "General transformation theory of differential elements," *American Journal of Mathematics*, vol. 32 (1904), pp. 392-401.

Any other transformation T converts either at most $\infty^{(n^2+5n+2)/2}$ or ∞^∞ unions into unions.

In a later section, we shall prove that all union-preserving transformations may be defined by a single directrix equation or a pair or triplet of such directrix equations.

4. The degenerate union-preserving transformations. A union-preserving transformation T is said to be nondegenerate if the rank of the jacobian-matrix

$$(8) \quad \begin{pmatrix} X_x & Y_x & Z_x \\ X_y & Y_y & Z_y \\ X_z & Y_z & Z_z \\ X_{p_j, m-j} & Y_{p_j, m-j} & Z_{p_j, m-j} \end{pmatrix},$$

formed from the first three columns of the jacobian-matrix (3), is three in a certain region of the (x, y, z) -space. If the rank is two or less our union-preserving transformation T is called degenerate. Thus a union-preserving transformation T is called degenerate if and only if there exists at least one functional relationship between the components (X, Y, Z) . We shall give a very brief discussion of the degenerate cases.

If all three functions (X, Y, Z) are identically constant, then T converts every union into a conical-union; these conical-unions all pass through a common point. If only two of the functions (X, Y, Z) are identically constant, then either $P = 0$ or $Q = 0$ or $1/P = 0$ or $1/Q = 0$, in which case every union is converted into a strip whose points describe a fixed line parallel to one of the coordinate axes. Finally if only one of the functions is identically constant, then either $P = Q = 0$, or $P/Q = 1/Q = 0$, or $Q/P = 1/P = 0$, in which case every union is converted into a plane parallel to one of the coordinate planes.

If there exist exactly two functional relationships between (X, Y, Z) , then every union is carried into a single fixed strip.

Finally if there is a single functional relationship between the three components (X, Y, Z) , then every union is carried into a single fixed surface.

Henceforth we shall consider only nondegenerate union-preserving transformations.

5. The general union-preserving transformations. We shall say that a non-degenerate union-preserving transformation T is general if it does not convert every conical-union of order n into a strip or conical-union. It is

said to be *intermediate* if every conical-union is carried into a strip. Finally we shall term T *special* if every conical-union is sent into a conical-union. Thus a non-degenerate transformation T of the union-preserving type is either general, or intermediate, or special.

A union-preserving transformation T is general, or intermediate, or special according as the rank of the jacobian-matrix

$$(9) \quad (X_{p_{j,n-j}}, \quad Y_{p_{j,n-j}}, \quad Z_{p_{j,n-j}}) \text{ for } j = 0, 1, 2, \dots, n,$$

is of rank two, or one, or zero in a certain region of the (x, y, z) -space. Of course, by Theorem 1, the rank of this matrix cannot be three.

In this section, we shall study general union-preserving transformations so that the rank of the jacobian-matrix (9) is two.

THEOREM 2. *For a general union-preserving transformation T , the $(n+1)$ independent variables $p_{j,n-j}$ for $j = 0, 1, 2, \dots, n$, can be eliminated from the three functions (X, Y, Z) ; thus obtaining the single eliminant*

$$(10) \quad \Omega(X, Y, Z, x, y, z, p_{10}, p_{01}, \dots, p_{n-1,0}, \dots, p_{0,n-1}) = 0.$$

We call this the *directrix equation* of our general union-preserving transformation T .

The above result is an immediate consequence of the first $(n+1)$ equations (7). For clearly the jacobian-matrix (9) is of rank two. Thus by the theory of jacobians, we find that all the partial derivatives of order n may be eliminated from the three functions (X, Y, Z) yielding the single directrix equation (10).

THEOREM 3. *A general union-preserving transformation is determined completely by a single directrix equation.*

Thus a general union-preserving transformation T depends on one arbitrary function of $(n^2 + n + 8)/2$ independent variables. The content of the set of general union-preserving transformations is $\infty^{1/[(n^2+n+8)/2]}$.

In the first place, it is obvious that under a general union-preserving transformation T , any point in the (X, Y, Z) -space corresponds to a family of ∞^∞ surfaces in the (x, y, z) -space, defined by a partial differential equation of order $(n-1)$. These surfaces are defined by the directrix equation (10) where (X, Y, Z) are considered as constants. The partial derivatives of these surfaces of orders n at most, satisfy the equations

$$(11) \quad \Omega = 0, \quad A_{n-1}(\Omega) = 0, \quad B_{n-1}(\Omega) = 0.$$

Secondly, any conical-union of order n in the (x, y, z) -space is converted

into a union in the (X, Y, Z) -space. This union is defined by the directrix equation (10) where $(x, y, z, p_{10}, p_{01}, \dots, p_{n-1,0}, \dots, p_{0,n-1})$ are regarded as constants. The partial derivatives of first order of this union satisfy the two equations

$$(12) \quad \Omega_X + P\Omega_Z = 0, \quad \Omega_Y + Q\Omega_Z = 0.$$

We shall prove that the functions (X, Y, Z, P, Q) of our general union-preserving transformation T must satisfy the equations (11) and (12). It is seen that the components (X, Y, Z) of T satisfy the directrix equation (10) since it was obtained as a result of eliminating the partial derivatives of order $n: p_{j,n-j}$ from the first three of equations (1) defining T .

Next let us differentiate the directrix equation (10) with respect to $p_{j,n-j}$. Upon using the first $(n+1)$ conditions of (7), we find that the result is

$$(13) \quad X_{p_{j,n-j}}(\Omega_X + P\Omega_Z) + Y_{p_{j,n-j}}(\Omega_Y + Q\Omega_Z) = 0.$$

Since for a general union-preserving transformation T , the rank of the jacobian-matrix (9) is two, it follows from (13) that we deduce the equations (12).

Apply the linear operators (5) where $k = n-1$ to the directrix equation (10) where (X, Y, Z) are given by the first three of equations (1). Because equations (12) are satisfied, we obtain the equations (11).

That we can actually solve (11) and (12) for (X, Y, Z, P, Q) follows from the fact that the directrix equation (10) represents a three-parameter family of partial differential equations of order $(n-1)$ in the (x, y, z) -space.

6. The intermediate union-preserving transformations. In this section, we shall consider the union-preserving transformations T whereby every conical-union of order n is converted into a strip. Thus the rank of the jacobian-matrix (9) is one.

THEOREM 4. *For an intermediate union-preserving transformation T , the $(n+1)$ independent variables $p_{j,n-j}$ for $j = 0, 1, 2, \dots, n$, can be eliminated from the three functions (X, Y, Z) , thus obtaining the two eliminants*

$$(14) \quad \begin{aligned} \Omega_1(X, Y, Z, x, y, z, p_{10}, p_{01}, \dots, p_{n-1,0}, \dots, p_{0,n-1}) &= 0, \\ \Omega_2(X, Y, Z, x, y, z, p_{10}, p_{01}, \dots, p_{n-1,0}, \dots, p_{0,n-1}) &= 0. \end{aligned}$$

These are the two directrix equations of any intermediate union-preserving transformation T .

The above result follows from the fact that the rank of the jacobian-matrix (9) is unity.

THEOREM 5. *The five functions (X, Y, Z, P, Q) defining an intermediate union-preserving transformation T satisfy not only the two directrix equations (14) but also the three equations*

$$(15) \quad \frac{\Omega_{1X} + P\Omega_{1Z}}{\Omega_{2X} + P\Omega_{2Z}} = \frac{\Omega_{1Y} + Q\Omega_{1Z}}{\Omega_{2Y} + Q\Omega_{2Z}} = \frac{A_{n-1}(\Omega_1)}{A_{n-1}(\Omega_2)} = \frac{B_{n-1}(\Omega_1)}{B_{n-1}(\Omega_2)}.$$

The transformation T is obtained by solving (14) and (15) for (X, Y, Z, P, Q) .

Thus a pair of directrix equations (14) determines completely an intermediate union-preserving transformation T . Any such transformation T is determined by two arbitrary functions of $(n^2 + n + 6)/2$ independent variables. The content of the set of intermediate union-preserving transformations is $\infty^{2f[(n^2+n+6)/2]}$.

To prove our Theorem 5, we proceed in the following manner. First of all, the components (X, Y, Z) satisfy the two directrix equations (14) since they were obtained as a result of eliminating the partial derivatives of order n : $p_{j,n-j}$ from the first three of equations (1) defining T . Moreover, these two directrix equations (14) are functionally independent.

Next differentiate each of the directrix equations (14) with respect to $p_{j,n-j}$. Introducing the first $(n+1)$ conditions (7), we obtain the results

$$(16) \quad \begin{aligned} X_{p_{j,n-j}}(\Omega_{1X} + P\Omega_{1Z}) + Y_{p_{j,n-j}}(\Omega_{1Y} + Q\Omega_{1Z}) &= 0, \\ X_{p_{j,n-j}}(\Omega_{2X} + P\Omega_{2Z}) + Y_{p_{j,n-j}}(\Omega_{2Y} + Q\Omega_{2Z}) &= 0. \end{aligned}$$

Since the rank of the jacobian-matrix (9) is unity, the ratios $X_{p_{j,n-j}}/Y_{p_{j,n-j}}$, are all equal for $j=0, 1, 2, \dots, n$. Also not all of the expressions $X_{p_{j,n-j}}$ and $Y_{p_{j,n-j}}$ are identically zero. For otherwise by (7), the rank of the jacobian-matrix (9) would be zero. Hence from these remarks, we find that the equations (16) yield only a single eliminant which is equivalent to the first proportion of (15).

Apply the linear operators (5) where $k=n-1$ to the directrix equations (14) where (X, Y, Z) are given by the first three of equations (1). Because of these conditions (7), we find

$$(17) \quad \begin{aligned} (\Omega_{1X} + P\Omega_{1Z})A_{n-1}(X) + (\Omega_{1Y} + Q\Omega_{1Z})A_{n-1}(Y) + A_{n-1}(\Omega_1) &= 0, \\ (\Omega_{2X} + P\Omega_{2Z})A_{n-1}(X) + (\Omega_{2Y} + Q\Omega_{2Z})A_{n-1}(Y) + A_{n-1}(\Omega_2) &= 0, \end{aligned}$$

and a similar set of two equations where the operator A_{n-1} is replaced by the operator B_{n-1} .

Because of the first proportion of (15) which was established in the earlier paragraphs, it follows that for the equations (17) and the similar set where the operator A_{n-1} is replaced by the operator B_{n-1} , to be valid, we must have the continued proportion (15).

It is assumed that Ω_1 and Ω_2 are so constructed that five of the equations derived from (14) and (15) are functionally independent with respect to (X, Y, Z, P, Q) . Hence these can be solved actually for (X, Y, Z, P, Q) ; thus obtaining our intermediate union-preserving transformation T .

7. The special union-preserving transformations. A union-preserving transformation is special if it converts every conical-union of order n into a conical-union of planar-elements. In this event, the components (X, Y, Z) will not involve the partial derivatives of order n : $p_{j,n-j}$, explicitly. Thus the first $(n+1)$ conditions of (7) are satisfied identically. The P and Q of such a transformation are determined by the remaining two equations of (7).

From the above remarks, it is seen that *a special union-preserving transformation is determined completely by a system of three functionally independent directrix equations of the forms (10).*

The special union-preserving transformations depend on three arbitrary functions of $(n^2 + n + 4)/2$ independent variables. The content of the set of special union-preserving transformations is $\infty^{3[f(n^2+n+4)/2]}$.

This completes our discussion of the union-preserving transformations. The non-degenerate union-preserving transformations are general, intermediate, or special. *These are determined by a single, pair, or triplet of directrix equations, respectively.* The contents of the three different classes are $\infty^{1f[(n^2+n+8)/2]}$, $\infty^{2f[(n^2+n+6)/2]}$, $\infty^{3f[(n^2+n+4)/2]}$, respectively.

It is noticed how the above result constitutes a direct extension of Lie's theorem on contact transformations.⁷

8. The union-preserving transformations in the domain of surface-elements. The following result concerning union-preserving transformations from surface-elements of order $n \geq 2$ into surface-elements of second order, will be established.

⁷ Kasner, "Lineal element transformations of space for which normal congruences of curves are converted into normal congruences," *Duke Mathematical Journal*, vol. 5 (1939), pp. 72-83. Also the following two papers by Kasner and De Ciccio: "Curvature element transformations which preserve integrable fields," *Proceedings of the National Academy of Sciences*, vol. 25 (1939), pp. 104-111; and "Transformation theory of integrable double-series of lineal elements," *Bulletin of the American Mathematical Society*, vol. 46 (1940), pp. 93-100.

THEOREM 6. *All the non-degenerate union-preserving transformations from surface-elements of order $n \geq 2$ into surface-elements of second order, are the extensions of the union-preserving transformations of surface-elements of order $(n-1)$ into planar-elements.*

A non-degenerate transformation from surface-elements of order $n \geq 2$: $(x, y, z; p_{j,m-j})$ for $j=0, 1, 2, \dots, m$ and $m=1, 2, \dots, n$, into surface-elements of second order: (X, Y, Z, P, Q, R, S, T) is given by equations of the forms

$$(18) \quad \begin{aligned} X &= X(x, y, z; p_{j,m-j}), & Y &= Y(x, y, z; p_{j,m-j}), & Z &= Z(x, y, z; p_{j,m-j}), \\ P &= P(x, y, z; p_{j,m-j}), & Q &= Q(x, y, z; p_{j,m-j}), \\ R &= R(x, y, z; p_{j,m-j}), & S &= S(x, y, z; p_{j,m-j}), & T &= T(x, y, z; p_{j,m-j}), \end{aligned}$$

where the jacobian-matrix of these eight functions is of rank eight in a certain region of the (x, y, z) -space.

We seek to find all those unions of surface-elements of order n which are converted into unions of surface-elements of second order by this transformation (18). The differential conditions (4) must correspond to the conditions

$$(19) \quad dZ = PdX + QdY, \quad dP = RdX + SdY, \quad dQ = SdX + TdY.$$

Thus by (5), we find the equations

$$(20) \quad \begin{aligned} A_n(Z) &= PA_n(X) + QA_n(Y), & B_n(Z) &= PB_n(X) + QB_n(Y), \\ A_n(P) &= RA_n(X) + SA_n(Y), & B_n(P) &= RB_n(X) + SB_n(Y), \\ A_n(Q) &= SA_n(X) + TA_n(Y), & B_n(Q) &= SB_n(X) + TB_n(Y), \end{aligned}$$

of order $(n+1)$. In general, there are at most $\infty^{(n^2+5n-6)/2}$ or ∞^∞ unions which become unions.

From the above equations, we discover that if the transformation (18) is union-preserving, it must satisfy the conditions

$$(21) \quad \begin{aligned} Z_{p_{j,n-j}} &= PX_{p_{j,n-j}} + QY_{p_{j,n-j}}, & A_{n-1}(Z) &= PA_{n-1}(X) + QA_{n-1}(Y), \\ & & B_{n-1}(Z) &= PB_{n-1}(X) + QB_{n-1}(Y), \\ P_{p_{j,n-j}} &= RX_{p_{j,n-j}} + SY_{p_{j,n-j}}, & A_{n-1}(P) &= RA_{n-1}(X) + SA_{n-1}(Y), \\ & & B_{n-1}(P) &= RB_{n-1}(X) + SB_{n-1}(Y), \\ Q_{p_{j,n-j}} &= SX_{p_{j,n-j}} + TY_{p_{j,n-j}}, & A_{n-1}(Q) &= SA_{n-1}(X) + TA_{n-1}(Y), \\ & & B_{n-1}(Q) &= SB_{n-1}(X) + TB_{n-1}(Y). \end{aligned}$$

Let us assume that not all the $X_{p_{j,n-j}}$ and $Y_{p_{j,n-j}}$ are zero. Accordingly our union-preserving transformation (18) is general or intermediate according

as the rank of the jacobian-matrix $(X_{p_{j,n-j}}, Y_{p_{j,n-j}})$ is two or one in a certain region of the (x, y, z) -space.

First consider the general case. From (X, Y, Z) of (18), we may eliminate all the partial derivatives of order n , obtaining a single eliminant of the form

$$(22) \quad Z = F(X, Y, x, y, z; p_{j,m-j})$$

where $j = 0, 1, 2, \dots, m$, and $m = 1, 2, \dots, n-1$.

Substituting this into the conditions (21), we find the following results

$$(23) \quad \begin{aligned} P &= F_X, & Q &= F_Y, & R &= F_{XX}, & S &= F_{XY}, & T &= F_{YY}, \\ A_{n-1}(F) &= 0, & B_{n-1}(F) &= 0, & A_{n-1}(F_X) &= 0, \\ B_{n-1}(F_X) &= 0, & A_{n-1}(F_Y) &= 0, & B_{n-1}(F_Y) &= 0. \end{aligned}$$

Thus the components (X, Y, Z, P, Q, R, S, T) of our transformation (18) must satisfy the equations (22) and (23).

Upon differentiating the conditions $A_{n-1}(F) = 0$, $B_{n-1}(F) = 0$, with respect to the $p_{j,n-j}$, we find, in view of the last four of equations (23), the conditions $F_{p_{j,n-j}} = 0$. Hence our transformation (18) must satisfy the conditions

$$(24) \quad A_{n-2}(F) = 0, \quad B_{n-2}(F) = 0, \quad F_{p_{j,n-1-j}} = 0.$$

At least one of the components (X, Y) must be effectively present in these equations. For otherwise these equations are identities and thus our eliminant (22) is of the form $Z = F(X, Y)$. This contradicts the fact that our transformation (18) is not degenerate.

Next it is seen that this system of equations (24) is not satisfied by a single solution of the form: $Y = Y(X, x, y, z; p_{j,m-j})$, (or $X = X(Y, x, y, z; p_{j,m-j})$), where $j = 0, 1, 2, \dots, m$, and $m = 1, 2, \dots, n-1$. For then our transformation (18) is defined by two eliminants and this contradicts the fact that the transformation (18) is general.

Thus it follows that the system (24) must have a common solution of the form: $X = X(x, y, z; p_{j,m-j})$, $Y = Y(x, y, z; p_{j,m-j})$, which involves partial derivatives of order $n-1$, at most. This contradicts the fact that our transformation (18) is general.

Therefore we have succeeded in establishing that there are no general union-preserving transformations from surface-elements of order $n \geq 2$ into surface-elements of second order.

Finally consider the intermediate case. From (X, Y, Z) of (18), we can eliminate the partial derivatives of order n , obtaining two functionally independent eliminants of the forms

$$(25) \quad Y = F(X, x, y, z; p_{j,n-j}), \quad Z = G(X, x, y, z; p_{j,m-j}),$$

for $j = 0, 1, 2, \dots, m$, and $m = 1, 2, \dots, n-1$. These are found on the assumption that at least one $X_{p_{j,n-j}}$ is not zero. Otherwise a $Y_{p_{j,n-j}}$ is not zero and the roles of X and Y are interchanged in these equations (25).

Substitute (25) into those conditions of (21) which involve Z . We find

$$(26) \quad G_X = P + QF_X, \quad A_{n-1}(G) = QA_{n-1}(F), \quad B_{n-1}(G) = QB_{n-1}(F).$$

It is observed that the equations $A_{n-1}(F) = 0$, and $B_{n-1}(F) = 0$ (and similarly when F is replaced by G) can not hold simultaneously. For otherwise we can differentiate these with respect to the $p_{j,n-j}$, and thereby deduce that $F_{p_{j,n-j}} = 0$. Thus the function X must satisfy the conditions

$$(27) \quad A_{n-2}(F) = 0, \quad B_{n-2}(F) = 0, \quad F_{p_{j,n-j}} = 0.$$

If X is not present in these equations, they must be identities. The first of equations (25) reduces to the form $Y = F(X)$. This contradicts the fact that our transformation (18) is not degenerate.

In the other possibility, the system (27) must possess a common solution X , which can involve partial derivatives of order $n-1$, at most. This contradicts our assumption that at least one $X_{p_{j,n-j}}$ is not zero.

The above arguments demonstrate that $A_{n-1}(F)$ and $B_{n-1}(F)$ (and similarly when F is replaced by G) are not simultaneously zero. From these facts and (26), we find that Q has the form

$$(28) \quad Q = H(X, x, y, z; p_{j,n-j}) = \frac{A_{n-1}(G)}{A_{n-1}(F)} = \frac{B_{n-1}(G)}{B_{n-1}(F)}.$$

Substitute H for Q into the equations of (21) which involve the partial derivatives of Q with respect to the $p_{j,n-j}$. We derive

$$(29) \quad (H_X - S - TF_X)X_{p_{j,n-j}} + H_{p_{j,n-j}} = 0.$$

These prove that the total partial derivatives of H with respect to the $p_{j,n-j}$ are proportional to the corresponding ones of X . Hence in the function H of (28), we must have $j = 0, 1, 2, \dots, m$, and $m = 1, 2, \dots, n-1$.

By using the conditions that the $Q_{p_{j,n-j}}$ are proportional to the $X_{p_{j,n-j}}$, we find by (28) that

$$(30) \quad Q = \frac{A_{n-2}(G)}{A_{n-2}(F)} = \frac{B_{n-2}(G)}{B_{n-2}(F)} = \frac{G_{p_{j,n-j}}}{F_{p_{j,n-j}}}.$$

If this system of equations possesses a common solution for X , then X can involve partial derivatives of order $(n-1)$, at most. This contradicts our assumption that at least one $X_{p_{j,n-j}}$ is not zero.

The only other remaining case to be considered is when the equations (30) are all identities. All the partial derivatives of G with respect to the small letters are proportional to the corresponding ones of F in this case. Therefore, all the small letters may be eliminated from the two equations (25), yielding an eliminant of the form: $\lambda(X, Y, Z) = 0$. This contradicts the fact that our transformation (18) is not degenerate.

Thus it has been established that there are no intermediate union-preserving transformations from surface-elements of order $n \geq 2$ into surface-elements of second order.

Therefore the only union-preserving transformations of the required kind are the extensions of the union-preserving transformations from surface-elements of order $(n - 1)$ into planar-elements. Theorem 6 is proved completely.

THEOREM 7. *The only union-preserving transformations from surface-elements of order $n \geq 1$ into surface-elements of order m where $n \geq m \geq 1$, are firstly, the contact group of planar-elements of Lie, and secondly, the union-preserving transformation from surface-elements of order n , where n is 2 or more, into planar-elements, together with the extensions of these two types.*

For, by hypothesis, any two unions which have contact of order $n \geq 2$ must be carried into two unions of second-order contact, at least. By Theorem 6, our correspondence must be consequently an extension of a union-preserving transformation from surface-elements of order $(n - 1)$ into planar-elements.

Therefore if T is a union-preserving transformation from surface-elements of order n into surface-elements of order m where $n \geq m \geq 1$, T is the extension of order m of a Lie contact transformation if $n = m$; and if $n > m$, T is the extension of order $(m - 1)$ of a union-preserving transformation from surface-elements of order $(n - m + 1)$ into planar-elements.

COLUMBIA UNIVERSITY,
NEW YORK, N. Y.

A NEW DEFINITION OF THE GODEAUX SEQUENCE OF QUADRICS.*

By CHENKUO PA.

A sequence of quadrics has been defined by L. Godeaux¹ at a generic point of an analytic non-developable surface in ordinary space. The first of them, ϕ , is the quadric of Lie and any two consecutive quadrics ϕ_{n-1} , ϕ_n in the sequence touch each other at four points, which are in turn the characteristic points of both quadrics. The second quadric ϕ_1 has been found independently by B. Su, who called it the associate quadric.² Godeaux³ has shown the equivalence of this quadric and the quadric ϕ_1 . In connection with these investigations the present author⁴ has generalized this quadric from the point of view of Finikoff.⁵ Following Godeaux, the definition of a Godeaux sequence consists in representing in S_5 the asymptotic tangents of a surface, and consequently has an indirect interpretation in S_3 . The object of this paper is two-fold; first, to give a geometrical significance of this sequence by means of certain linear complexes associated with a point of a surface, and second, to generalize this new definition to a W congruence so as to obtain a generalized sequence of quadrics.⁶

1. For the subsequent discussion, it is convenient to utilize the normal coordinate system of Cartan⁷ at a generic point M of a surface. Let $\{M, M_1, M_2, M_3\}$ be a tetrahedron of reference of such a system; we have, on putting

* Received August 24, 1945.

¹ L. Godeaux, "Sur les lignes asymptotiques d'une surface et l'espace réglée," *Bulletin de l'Académie royale de Belgique* (1930), pp. 812-826; (1928), pp. 31-41, or "La théorie des surfaces et l'espace réglé" (1934).

² Buchin Su, "On the surfaces whose asymptotic curves belong to linear complexes II," *Tôhoku Mathematical Journal*, vol. 40 (1935), pp. 433-448.

³ L. Godeaux, "Remarques sur les quadriques associées aux points d'une surface," *Journal of the Chinese Mathematical Society*, vol. 2 (1937), pp. 1-5.

⁴ Chenkuo Pa, "A generalization of associate quadrics of a surface," *American Journal of Mathematics*, vol. 66 (1944), pp. 115-121.

⁵ S. Finikoff, "Sur les quadriques de Lie et les congruences de M. Demoulin," *Recueil Mathématique de Moscou*, vol. 37 (1930), pp. 48-97.

⁶ Cf. L. Godeaux, "Sur une propriété de l'enveloppe de certaines familles de quadriques," *Rendiconti dei Lincei*, Ser. 6, vol. 11 (1930), pp. 54-58.

⁷ Cf. S. Finikoff, *Comptes Rendus*, vol. 197 (1933), pp. 883-885; B. Su, *Tôhoku Mathematical Journal*, vol. 41 (1935), pp. 203-215.

$$| M M_1 | = U, \quad | M M_2 | = V,$$

that U (or V) is the image of the asymptotic u -tangent (v -tangent) in the Klein hyperquadric Ω of S_5 . The locus of U (or V) forms a conjugate net on Ω and U satisfies the following Laplace equation:

$$\frac{\partial^2 U}{\partial u \partial v} - \frac{\partial \log \beta}{\partial v} \frac{\partial U}{\partial v} - \beta \gamma U = 0.$$

Thus it gives rise to a Laplace sequence self-conjugate to Ω , namely,

$$\cdots (U_{-n}), \cdots, (U_{-1}), (U), (U_1), \cdots, (U_n), \cdots,$$

where $U_{-1} = V$ and (U_n) denotes the Laplace transform of (U_{n-1}) in the sense v . Let h_i, k_i be the Darboux-Laplace invariants of the net (U_i) ; then we have the following relations:

$$\frac{\partial U_n}{\partial v} = U_{n+1} + U_n \frac{\partial}{\partial v} \log (\beta h_1 \cdots h_n),$$

$$\frac{\partial U_n}{\partial u} = h_n U_{n-1}, \quad (U_0 = U, \quad U_{-1} = V, \quad h_0 \equiv \beta),$$

$$\frac{\partial^2 U_n}{\partial u \partial v} - \frac{\partial U_n}{\partial v} \frac{\partial}{\partial v} \log (\beta h_1 \cdots h_n) - h_n U_n = 0;$$

$$\frac{\partial V_n}{\partial u} = V_{n+1} + V_n \frac{\partial}{\partial u} \log (\gamma k_1 \cdots k_n),$$

$$\frac{\partial V_n}{\partial v} = k_n V_{n-1}, \quad (V_0 = V, \quad V_{-1} = U, \quad k_0 \equiv \gamma),$$

$$\frac{\partial^2 V_n}{\partial u \partial v} - \frac{\partial V_n}{\partial u} \frac{\partial}{\partial u} \log (\gamma k_1 \cdots k_n) - k_n V_n = 0.$$

The intersection of the plane $(U_n U_{n+1} U_{n+2})$ with Ω is known to be the image in S_5 of one regulus of the quadric ϕ_n , namely, the n -th quadric in the Godeaux sequence.

Let us now consider the osculating linear complex $R_1(u, v)$ of the asymptotic u -curve at the point $M(u, v)$. It can easily be shown that this complex is represented in S_5 by the intersection of the hyperplane $[U, V, V_1, V_2, V_3]$ with Ω . If $M(u + du, v + dv)$ be any point of S near $M(u, v)$, the corresponding linear complex $R_1(u + du, v + dv)$ may then be expressed by $R_1 + R_{1u} du + R_{1v} dv + \cdots$ and therefore all the linear complexes corresponding to points in the first order neighborhood of M have a common regulus formed by lines common to $R_1(u, v)$, $R_{1u}(u, v)$, $R_{1v}(u, v)$. This regulus constitutes the quadric ϕ of Lie, as was shown in the former paper.

We now consider the linear complexes $R_1(u, v + dv) = R_1 + R_{1v}dv + (1/2!)R_{1vv}dv^2 + \dots$ corresponding to points in the second order neighborhood of M along an asymptotic v -curve. We find easily that these linear complexes have one regulus of the quadric ϕ_1 in common, and the linear complexes $R_1(u, v + dv) = R_1 + R_{1u}du + (1/2!)R_{1vu}dv^2 + (1/3!)R_{1vuu}dv^3 + \dots$ corresponding to points in the third order neighborhood of M along an asymptotic v -curve have two common straight lines g_1, g_2 , whose images on Ω are the points of intersection of (V_2V_3) with Ω . When these lines g_1, g_2 vary in the direction u , we obtain two ruled surfaces G_1, G_2 . There exists a linear complex $R_2(u, v)$ having a contact of the third order with both G_1 and G_2 along the generators g_1, g_2 respectively. In fact this complex $R_2(u, v) \equiv R_2$ is determined by $[V_2V_3V_4V_5V_6]$. With the aid of $R_2(u, v)$, we obtain, on making use of a similar method, the quadrics ϕ_2, ϕ_3, ϕ_4 . More precisely, they are determined as common reguli of the three sets of linear complexes $R_2 + R_{2u}du + (1/2!)R_{2uu}du^2$, $R_2 + R_{2u}du + R_{2v}dv$ and $R_2 + R_{2v}dv + (1/2!)R_{2vv}dv^2$ respectively. If we consider all the linear complexes $R_2 + R_{2v}dv + (1/2!)R_{2vv}dv^2 + (1/3!)R_{2vvr}dv^3$ corresponding to points in the third order neighborhood along an asymptotic v -curve, we get two straight lines similar to g_1, g_2 . Applying the process successively we can define the whole sequence.

Thus we have proved the following

THEOREM. *The Godeaux sequence of quadrics associated with a point of a surface can be simply defined with the aid of asymptotic osculating linear complexes.*

2. In the next place let us consider the asymptotic osculating ruled surface R_u generated by v -tangents along an asymptotic u -curve. The image of the osculating linear complex $S_1(u, v)$ of R_u at M is evidently the hyperplane $[VV_1V_2V_3V_4]$. All the linear complexes $S_1 + S_{1u}du + (1/2!)S_{1uu}du^2$ corresponding to points of the surface in the second order neighborhood of M along an asymptotic u -curve have one regulus of the quadric ϕ in common; all the linear complexes $S_1 + S_{1u}du + S_{1v}dv$ corresponding to points of the surface in the first order neighborhood of M have one regulus of the quadric ϕ_1 in common, and all the linear complexes $S_1 + S_{1v}dv + (1/2!)S_{1vv}dv^2$ corresponding to points of the surface in the second order neighborhood of M along an asymptotic v -curve have their common lines in one regulus of the quadric ϕ_2 . Similarly, among the linear complexes $S_1 + S_{1u}du + (1/2!)S_{1vu}dv^2 + (1/3!)S_{1vuu}dv^3$ corresponding to points of the surface in the third order neighborhood of M along an asymptotic v -curve, there are only four inde-

pendent ones, namely, $S_1, S_{1v}, S_{1vv}, S_{1rvv}$ and, in consequence, *they have two common straight lines*. Proceeding in the same manner as in **1**, we can define the quadrics ϕ_3, ϕ_4, ϕ_5 successively, and thus obtain the whole sequence. In summary, we can state the following

THEOREM. *With the aid of osculating linear complexes of the asymptotic ruled surfaces associated with a point of a surface, we can also define the Godeaux sequence of quadrics.*

3. There is now no difficulty in generalizing the notion of a Godeaux sequence of quadrics to a W congruence.⁸ For, by a classical theorem of Darboux, the image (U) of a W congruence in the Klein quadric of S_5 forms a conjugate net. Letting (U_{i+1}) be the Laplace transform of (U_i) in the sense v , we get a Laplace sequence

$$\cdots, (U_{-n-1}), (U_{-n}), \cdots, (U_{-1}), (U), (U_1), \cdots, (U_n), (U_{n+1}), \cdots$$

Hence by the method used by Godeaux in the case of the asymptotic congruence, we can define a sequence of quadrics:

$$\cdots, \phi_{-n}, \cdots, \phi_{-2}, \phi_{-1}, \phi, \phi_1, \phi_2, \cdots, \phi_n, \cdots,$$

where ϕ_n is defined by $(U_n U_{n+1} U_{n+2})$.

It should be mentioned that the quadric ϕ_{-1} decomposes into two focal planes of the congruence, and this sequence has properties similar to the Godeaux sequence.

By the method used in **1** and **2**, we can readily define the new sequence. In the present case, the quadrics of the sequence can, however, be determined by the *osculating complexes* of the congruence.

In fact, the osculating linear complex of a W congruence is determined by $(U_{-2} U_{-1} U U_1 U_2)$. By the method of **2**, we first define the quadrics $\phi_{-2}, \phi_{-1}, \phi$, and then obtain, on using the same complex in the sense v , the quadrics ϕ_1, ϕ_2, ϕ_3 , and similarly, the quadrics $\phi_{-3}, \phi_{-4}, \phi_{-5}$, etc.

Thus we have arrived at the following

THEOREM. *With the aid of osculating complexes of a W congruence, a sequence of quadrics can be defined to be associated with each ray of the congruence, that is, the generalization of a Godeaux sequence of a surface to a W congruence.*

NATIONAL UNIVERSITY OF CHEKIANG,
KWEICHOW, MEITAN, CHINA.

⁸ L. Godeaux, *loc. cit.* ⁸.

UNILATERAL VARIATIONS WITH VARIABLE END-POINTS.*

By JULIAN D. MANCILL.

1. **Introduction.** For the derivation of certain necessary conditions for an arc E_{12} to minimize an integral in the form

$$I = \int_{t_1}^{t_2} F(x, y, x', y') dt \equiv \int_{t_1}^{t_2} F(x, x') dt,$$

we shall assume that $F(x, x')$, $F_{x'}(x, x')$, and $F_{y'}(x, x')$ are continuous in a region R of the xy -plane, and for all $(x', y') \neq (0, 0)$. The function $F(x, x')$ is assumed to be positively homogenous of the first degree in x' and y' . Admissible curves will be supposed always in the form

$$x = x(t), \quad y = y(t),$$

and of class D' in the region R .¹ In order to emphasize the problem at hand, *we shall assume, unless otherwise stated, that the minimizing curve E_{12} lies along the boundary of the region R of admissible curves.* Our problem then is to derive the properties of the arc E_{12} in order that it minimize the integral I in the class of all admissible curves joining a given curve D through the point 1 and a given curve C through the point 2. We shall assume that the curves D and C are of class C' in the region R .

The usual procedure in deriving necessary conditions in order that an arc minimize the integral I in the case of free variations has been to derive the Euler equations first² and then to make use of them in deriving further necessary conditions. Recent studies have shown that such a procedure is not necessary nor desirable for the derivation of two of the so called first order conditions. These two conditions for the arc E_{12} are the Weierstrass condition,³ which states that

$$\mathcal{E}(x, x', p) \equiv F(x, p) - pF_{x'}(x, x') - qF_{y'}(x, x') \geq 0$$

along E_{12} for all $(p, q) \neq (0, 0)$, and the corner conditions.⁴

* Received October 22, 1945; Revised September 5, 1946.

¹ For a definition of the term "class" as here used see Bolza, *Lectures on the calculus of variations*, University of Chicago, p. 7.

² See, e. g., Bolza, *loc. cit.*, p. 122.

³ Graves, "A proof of the Weierstrass condition in the calculus of variations," *American Mathematical Monthly*, vol. 41 (1934), pp. 502-504.

⁴ Mancill, "A proof of the corner conditions in the calculus of variations," *American Mathematical Monthly*, vol. 43 (1936), pp. 68-70.

In the present paper we shall derive a third first order condition, the transversality condition at the points 1 and 2, without making use of the Euler equations. These transversality conditions involve inequalities in the case of unilateral variations instead of equalities as in the case of free variations. The proof is made with weaker hypotheses than the usual proof by means of the first variation. The method of proof is simple and may be applied to parametric or non-parametric problems in space of any number of dimensions. So far as the author knows these results are new since there seems to be no reference in the literature to transversality at points of the boundary of the region of admissible curves.

Jacobi's criterion will be formulated for the cases in which the Euler equations are satisfied along E_{12} . There is no restriction on the length of E_{12} in the case where the Euler equations are not satisfied along E_{12} .

A set of sufficient conditions in order that the arc E_{12} shall minimize the integral I will be given after appropriate modification of the original hypotheses on the region R and the function $F(x, x')$. There are many cases and they lead to a large number of distinct sets of sufficient conditions. The only cases not treated fully are the ones usually omitted, that is when $F(x, x') = 0$ along E_{12} at the points 1 and 2 when the curves D and C are transversal or tangent to E_{12} . The methods used are applicable to non-parametric problems.

The treatment of sufficient conditions given here for the case where E_{12} is an extremal and the curves D and C are transversal to E_{12} applies to free variations, and the method of proof is simpler than that given by Bliss⁵ for non-parametric problems. He assumes that the integrand function is positive throughout the region of admissible curves for all directions, whereas in the present treatment it is only assumed that $F(x, x') \neq 0$ along E_{12} at the points 1 and 2.

2. The transversality condition. Let the equations of the minimizing curve E_{12} be

$$\bar{x} = x(t), \quad y = y(t), \quad t_1 \leq t \leq t_2,$$

and those of the curves D and C be respectively

$$D: \quad x = X_1(\tau), \quad y = Y_1(\tau),$$

and

$$C: \quad x = X_2(\tau), \quad y = Y_2(\tau),$$

⁵ "Jacobi's criterion when both end-points are variable," *Mathematische Annalen*, vol. 58 (1904), pp. 70-80.

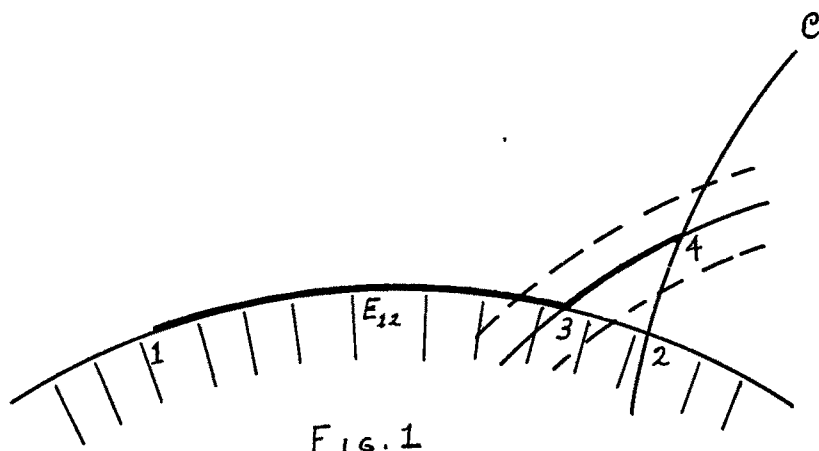
interior to the region R . Let the equations

$$\mathcal{E}: \quad x = \phi(s, a), \quad y = \psi(s, a)$$

represent a one parameter family of admissible curves intersecting E_{12} for $s = s_3(a)$, $t = t(a)$, and intersecting the curve C for $s = s_4(a)$, $\tau = \tau(a)$ in a neighborhood of the point 2. The incidence relations of these curves are expressed by the equations

$$(1) \quad \begin{aligned} x[t(a)] &= \phi[s_3(a), a], & y[t(a)] &= \psi[s_3(a), a], \\ X_2[\tau(a)] &= \phi[s_4(a), a] & Y_2[\tau(a)] &= \psi[s_4(a), a], \end{aligned}$$

where we may assume $t'(a) > 0$ at $a = a_2$.



Now consider the admissible comparison curve

$$E_a \equiv E_{12} + \mathcal{E}_{34},$$

of the type indicated by the heavy line in the figure. The function

$$I(a) = I(E_a)$$

has the derivative

$$(2) \quad I'(a_2) = F(x, x')t'(a_2) - F(x, \phi')s'_3(a_2) + F(x, \phi')s'_4(a_2)$$

at $a = a_2$. By means of the homogeneity property of F , the incidence relations (1), and their derivatives with respect to a , we find that

$$(3) \quad I'(a_2) = \mathcal{E}(x, \phi', x')t'(a_2) + [X'_2 F_{x'}(x, \phi') + Y'_2 F_{y'}(x, \phi')] \tau'(a_2) \\ \leq 0$$

at the point 2 since E_{12} is to minimize the integral I and $t'(a_2)$ is assumed positive. The parameter t may be chosen so that $t'(a_2) = 1$. Now, if we let the direction (ϕ', ψ') approach the direction (x', y') at 2 and assume that C is not tangent to E_{12} at 2, it follows from the inequality (3) that

$$(4)_2 \quad X'_2 F_{x'}(x, x') + Y'_2 F_{y'}(x, x') \geq 0$$

if $\tau'(a_2) < 0$, that is if the curve C is interior to R for $\tau \geq \tau_2$. Likewise if $\tau'(a_2) > 0$, that is if C is interior to R for $\tau \leq \tau_2$, it follows from (3) that

$$(5)_2 \quad X'_2 F_{x'}(x, x') + Y'_2 F_{y'}(x, x') \leq 0.$$

The two inequalities $(4)_2$ and $(5)_2$ can be combined into the single condition

$$\bar{T}(x, x', X'_2) = [X'_2 F_{x'}(x, x') + Y'_2 F_{y'}(x, x')]/D(x', X'_2) \geq 0,$$

if

$$D(x', X'_2) = \begin{vmatrix} x' & X'_2 \\ y' & Y'_2 \end{vmatrix} \neq 0,$$

as the transversality condition at the variable end-point 2. This is true because $D(x', X'_2)$ is positive when $(4)_2$ is necessary and is negative when $(5)_2$ is necessary. This condition can be derived directly from the inequality (3) if we make the additional assumption that the well known function $F_1(x, x') > 0$ along E_{12} at the point 2. For then it follows that

$$[X'_2 F_{x'}(x, x') + Y'_2 F_{y'}(x, x')] \tau'(a_2) / t'(a_2) < 0$$

at the point 2 for the direction (ϕ', ψ') sufficiently near but distinct from the direction (x', y') of E_{12} at the point 2, since $\mathcal{E}(x, \phi', x') > 0$ at 2 under these conditions. From the derivatives of the relations (1), we have

$$\tau'(a_2) / t'(a_2) = D(\phi', x') / D(\phi', X'_2).$$

Since the admissible region R is to the left along E_{12} , the determinant $D(\phi', x')$ is negative and we have

^o Cf. Bolza, *loc. cit.*, p. 140.

$$[X'_2 F_{x'}(x, \phi') + Y'_2 F_{y'}(x, \phi')]/D(\phi', X'_2) > 0$$

at the point 2 if the direction (ϕ', ψ') is sufficiently near but distinct from the direction (x', y') there. Therefore, it follows that $\tilde{T}(x, x', X'_2) \geq 0$ if $D(x', X'_2) \neq 0$ at 2.

If $D(x', X'_2) = 0$ at the point 2, that is if $(X'_2, Y'_2) = (kx', ky')$ for $k \neq 0$, several cases arise:

1) If the curve C precedes the normal line to E_{12} at 2, then $s'_3(a_2) = s'_4(a_2)$ in (2) as follows from the derivatives of the relations (1). Therefore, in this case we have

$$F(x, x') \leq 0,$$

since we assumed that $t'(a_2) > 0$. This result holds for k positive or negative and is independent of the element (x, ϕ') used in the construction.

2) If the curve C lies beyond the normal line to E_{12} at 2, then we may make use of the family of admissible curves (\mathfrak{E}) following 2 and in the relation analogous to (2) we would have again $s'_3(a_2) = s'_4(a_2)$. Therefore, in this case

$$I'(a_2)/t'(a_2) = F(x, x') \geq 0$$

at the point 2. This result is true for k positive or negative and is independent of the element (x, ϕ') used in the construction.

3) If the curve C interior to R lies on both sides of the normal line to E_{12} at 2, that is if C does not cross E_{12} , then the results of cases 1) and 2) must hold simultaneously and therefore in this case

$$F(x, x') = 0.$$

Analogous conditions can be shown to hold at the point 1 between the curves E_{12} and D with each inequality reversed.

If either of the points 1 or 2 is interior to the region R of admissible curves, then a construction similar to the one used above but on the opposite side of E_{12} shows, with the results already obtained, that for free variations all inequalities in the conditions are to be replaced by equalities. That is, at an interior variable end-point

$$\tilde{T}(x, x', X') = 0,$$

if the transversal direction (X', Y') is distinct from the direction (x', y') , and

$$F(x, x') = 0$$

if $(X', Y') = (kx', ky')$ for $k \neq 0$.

We may summarize the results of this section in the following:

THEOREM 1. *If the curve E_{12} minimizes the integral I in the class of all admissible curves joining the curves D and C ; and*

1) *if the point 2 is interior to R , then $T(x, x'; X'_2) = 0$ for $(X'_2, Y'_2) \neq (kx', ky')$ ($k > 0$) and $F(x, x') = 0$ at 2 for $(X'_2, Y'_2) = (kx', ky')$ ($k \neq 0$);*

2) *if the arc E_{12} is along the boundary of R , then $T(x, x'; X'_2) \geq 0$ at 2 for $(X'_2, Y'_2) \neq (kx', ky')$ ($k > 0$) and for $(X'_2, Y'_2) = (kx', ky')$ ($k \neq 0$)*

$F(x, x') \leq 0$ if C precedes the normal line to E_{12} at 2,

$F(x, x') \geq 0$ if C lies beyond the normal to E_{12} at 2,

$F(x, x') = 0$ if C lies on both sides of the normal to E_{12} at 2.

Similar results hold at the point 1 for E_{12} and the curve D but with the inequalities reversed.

The conditions in the second part of Theorem 1 have a simple geometric interpretation. Suppose that the curve C is interior to R for $\tau \geq \tau_2$. Then the inequality $(4)_2$ must hold even when $(X'_2, Y'_2) = (kx', ky')$ ($k \neq 0$) since in this case

$$X'_2 F_{x'}(x, x') + Y'_2 F_{y'}(x, x') = kF(x, x') \geq 0$$

at 2. Similarly, if the curve D is interior to R for $\tau \geq \tau_1$, we have

$$(4)_1 \quad X'_1 F_{x'}(x, x') + Y'_1 F_{y'}(x, x') \leq 0$$

at the point 1. Consider now the two relations $(4)_1$ and $(4)_2$ along with the equations

$$(6) \quad x' F_{x'}(x, x') + y' F_{y'}(x, x') = F(x, x'),$$

and

$$(7) \quad X' F_{x'}(x, x') + Y' F_{y'}(x, x') = 0,$$

where (X', Y') denotes the direction transversal to (x', y') . The equations (6) and (7) in general involve three directions at a point of E_{12} , namely

(x', y') , (X', Y') , and (F_x, F_y) , where we assume for the moment that not both the derivatives F_x and F_y are zero. Equation (7) shows that the direction (X', Y') is orthogonal to (F_x, F_y) . Equation (6) shows that the angle between the directions (x', y') and (F_x, F_y) is acute if $F(x, x')$ is positive and is obtuse if $F(x, x')$ is negative. The inequality (4)₂ shows that the angle at 2 formed by the directions (X'_2, Y'_2) and (F_x, F_y) is acute, while

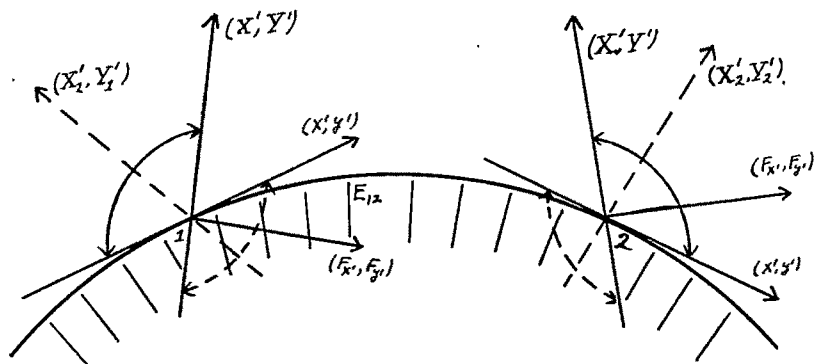


FIG. 2

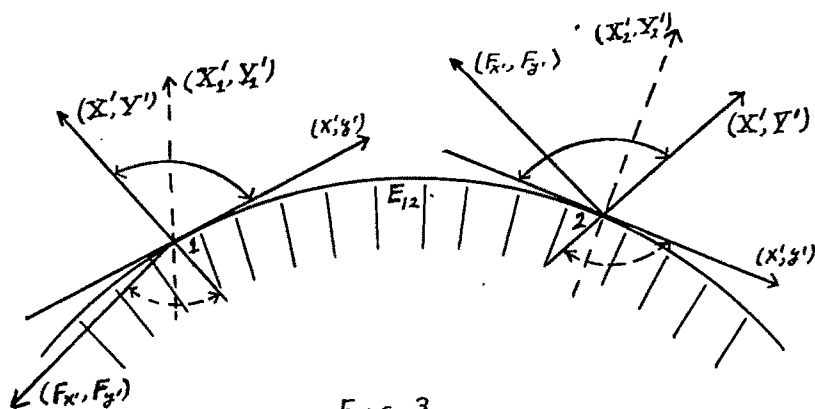


FIG. 3

the relations (4)₁ shows that the angle at 1 formed by the directions (X'_1, Y'_1) and (F_x, F_y) is obtuse. These relations are shown in figure 2 when F is positive at 1 and 2 and in figure 3 when F is negative at the points 1 and 2. The curved arrow indicates the possible range of the directions of the curves D and C directed toward the interior of R , and analogously the broken curved arrow indicates the range of the directions of D and C directed toward the exterior of R in each case. It follows, therefore that if $F(x, x') \neq 0$ at 1 and 2, then there exist directions for which the conditions of the Theorem are satisfied.

If $F(x, x') = 0$ at 1 and 2, then equations (6) and (7) show that the directions (x', y') and (X', Y') are coincident or in opposite directions at 1 or 2, unless $F_{x'}$ and $F_{y'}$ are both zero. In the latter case any direction may be defined as transversal to (x', y') . If $F_{x'}$ and $F_{y'}$ are not both zero but $F(x, x')$ is zero, there may or may not exist directions distinct from (x', y') for which the conditions of the Theorem are satisfied.

3. Sufficient conditions. For the treatment of sufficient conditions, as well as for Jacobi's criterion, we shall assume that $F(x, x')$ is of class C''' in a region R' containing R in its interior and for all $(x', y') \neq (0, 0)$. We shall assume that our problem is regular, that is $F_1(x, x') > 0$ in R for all $(x', y') \neq (0, 0)$. We shall suppose also that the curves D and C are of class C'' in R .

In addition to the function $\tilde{T}(x, x', X')$ already defined, we shall need the function⁷

$$T(x, x', x'') \equiv F_{x''}(x, x') - F_{y''}(x, x') + F_1(x, x')(x'y'' - x''y').$$

THEOREM 2. Suppose that E_{12} has the properties:

- 1) E_{12} is of class C'' and has $T < 0$ along it;
- 2) if $D(x', X'_1) \neq 0$ at 1, then $\tilde{T}(x, x', X'_1) \leq 0$ there; or if $D(x', X'_1) = 0$ at 1, then

$$\begin{aligned} F(x, x') &> 0 \text{ at 1 if } D \text{ precedes the normal to } E_{12} \text{ at 1,} \\ F(x, x') &< 0 \text{ at 1 if } D \text{ lies beyond the normal to } E_{12} \text{ at 1;} \end{aligned}$$

- 3) if $D(x', X'_2) \neq 0$ at 2, then $\tilde{T}(x, x', X'_2) \geq 0$ there; or if $D(x', X'_2) = 0$ at 2, then

$$\begin{aligned} F(x, x') &< 0 \text{ at 2 if } C \text{ precedes the normal to } E_{12} \text{ at 2,} \\ F(x, x') &> 0 \text{ at 2 if } C \text{ lies beyond the normal to } E_{12} \text{ at 2.} \end{aligned}$$

Then the arc E_{12} minimizes the integral I in the class of all admissible curves sufficiently near E_{12} joining the curves D and C .

Since $T < 0$ along E_{12} it remains less than zero along the boundary of R in neighborhoods preceding the point 1 and following the point 2. We shall

⁷ If E_{12} is of class C'' and minimizes the integral I , then $T \leq 0$ along it. For a proof see, Bolza, *loc. cit.*, p. 148.

denote this extended arc by E_{56} , where the points 5 and 6 are sufficiently near 1 and 2 respectively. Then the one parameter family of extremals

$$\mathcal{E}: x = \phi(t, a), \quad y = \psi(t, a),$$

tangent to E_{56} for $t = t(a)$, $a_5 \leq a \leq a_6$, forms a field \mathcal{F}_2 adjoining E_{56} in R for

$$t(a) \leq t \leq t(a) + e_2, \quad a_5 \leq t \leq a_6,$$

if e_2 is sufficiently small.⁸ Also, the family of extremals (\mathcal{E}) forms a field \mathcal{F}_1 adjoining E_{56} in R for

$$t(a) - e_1 \leq t \leq t(a), \quad a_5 \leq a \leq a_6,$$

if e_1 is sufficiently small. The functional determinant

$$\Delta(t, a) \equiv \begin{vmatrix} \phi' & \phi_a \\ \psi' & \psi_a \end{vmatrix}$$

is positive interior to \mathcal{F}_1 , negative interior to \mathcal{F}_2 , and is zero on E_{56} . Now choose a point 8 on E_{12} at which $F(x, x') \neq 0$. Let L represent a curve of class C' through the point 8 in the unique direction transversal to E_{12} there. The one parameter family of extremals

$$\tilde{\mathcal{E}}: x = \tilde{\phi}(t, a), \quad y = \tilde{\psi}(t, a),$$

cut by L transversally forms a field \mathcal{F}_3 in R for

$$t_8 - e_3 \leq t \leq t_8 + e_3, \quad a_8 \leq a \leq a_8 + d,$$

where e_3 and d are sufficiently small positive numbers. This family of extremals contains the extremal \mathcal{E}_8 of the family (\mathcal{E}) for $a = a_8$. Let the region in the xy -plane common to the regions \mathcal{F}_1 and \mathcal{F}_2 be denoted by \mathcal{F} . Let $V_{1'2'}$ denote any admissible comparison curve in \mathcal{F} joining the curves D and C in the points $1'$ and $2'$, respectively, and distinct from E_{12} . Let 7 represent the last point along $V_{1'2'}$, in which $V_{1'2'}$ crosses \mathcal{E}_8 preceding 8. Also, let 9 denote the first point beyond 7 along $V_{1'2'}$ in which $V_{1'2'}$ crosses \mathcal{E}_8 beyond 8. Then it is easily seen that \mathcal{F} can be so restricted that V_{79} lies in the region \mathcal{F}_3 .

Now, by considering only the field \mathcal{F}_1 and making use of the assumption of regularity, it can be shown that⁹

⁸ Bolza, *Vorlesungen über Variationsrechnung*, Berlin, 1909, pp. 403-05.

⁹ Mancill, "The minimum of a definite integral with respect to unilateral varia-

$$(8) \quad I(V_{1'7} + \mathcal{E}_{78}) > I(\mathcal{E}_{1'3} + E_{38}).$$

By considering the field \mathcal{F}_2 it can be shown in the same way that

$$(9) \quad I(\mathcal{E}_{89} + V_{92'}) > I(E_{84} + \mathcal{E}_{42'}).$$

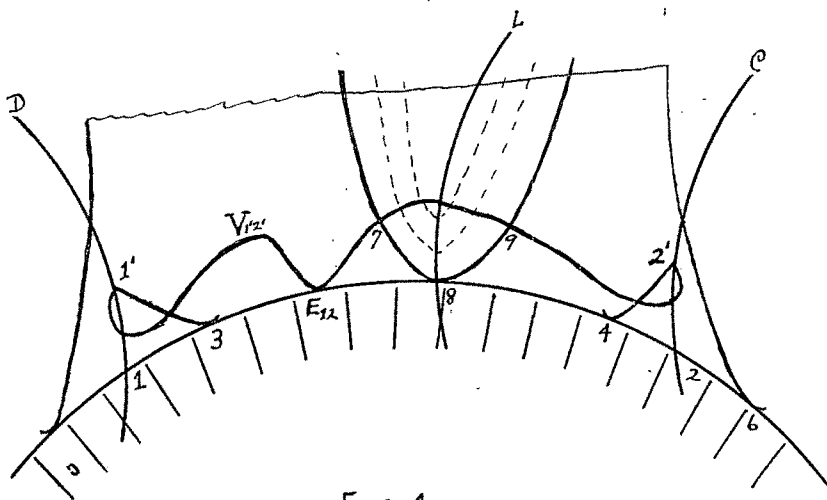


FIG. 4

Also, by making use of the well-known invariant integral I^* in the field \mathcal{F}_3 , it can be shown that

$$(10) \quad I(V_{79}) > I(\mathcal{E}_{78} + \mathcal{E}_{89}).$$

Combining the inequalities (8), (9), and (10), we obtain

$$(11) \quad I(V_{1'2'}) > I(\mathcal{E}_{1'3} + E_{34} + \mathcal{E}_{42'}).$$

It is our purpose now to show that in general, under the hypotheses of our theorem,

$$(12) \quad I(\mathcal{E}_{1'3}) > I(E_{13}),$$

and

$$(13) \quad I(\mathcal{E}_{42'}) > I(E_{42}),$$

if the field \mathcal{F} is sufficiently restricted.

tions," *Contributions to the Calculus of Variations*, 1933-1937, University of Chicago Press, pp. 129-32.

We shall prove the inequality (13) in two parts by making use of the extremals of the field \mathcal{F}_2 . First consider the case where $D(x', X'_2) \neq 0$ at the point 2. The function

$$I_2(a) \equiv I(E_{44'} + \mathcal{E}_{4'2''}),$$

where $4'$ is on E_{42} and $2''$ is on C , has the derivative

$$I'_2(a) = \Delta(t, a) \tilde{T}(x, \phi', X'_2),$$

which is negative near the point 2 and is zero at 2, if $\tilde{T}(x, x', X'_2) > 0$ there. If $\tilde{T}(x, x', X'_2) = 0$ at 2 we shall need to consider the derivative of the function

$$g_2(a) \equiv [X'_2 F_{x'}(x, \phi') + Y'_2 F_{y'}(x, \phi')]^{2''}$$

near 2. Direct calculation shows that this function has the derivative

$$g'_2(a) = [A_2 \Delta(t, a) + B_2 \Delta_t(t, a)] / D(\phi', X'_2)$$

near 2, where A_2 is a known function of the elements of our problem and $B_2 = [D(\phi', X'_2)]^2 F_1(x, \phi')$. Therefore, $g'_2(a)$ has the same sign near 2 as has the function

$$D(x', X'_2) F_1(x, x') \Delta_t(t_2, a_2)$$

at 2; this function is not zero and has the sign opposite to that of $D(x', X'_2)$ since $\Delta_t(t_2, a_2)$ is negative. Therefore, it follows that even if $\tilde{T}(x, \phi', X'_2) = 0$ at 2 it is positive near 2 in the field \mathcal{F}_2 , if this field is sufficiently restricted. Thus, $I'_2(a)$ is negative near 2 and the function $I_2(a)$ is decreasing. But it is obvious that

$$I(\mathcal{E}_{42'}) - I(E_{42}) = I_2(a_4) - I_2(a_2),$$

where $a_4 < a_2$ if $2'$ is distinct from 2, and the inequality (13) is proved.

Next we shall consider the case where $D(x', X'_2) = 0$, that is $(X'_2, Y'_2) = (kx', ky')$ ($k \neq 0$), at 2. Suppose that the curve C is interior to the region R for $\tau \geq \tau_2$. Then it follows from the hypotheses of the Theorem that

$$X'_2 F_{x'}(x, x') + Y'_2 F_{y'}(x, x') > 0$$

at 2. Consider the function ¹⁰

$$W(x, y) \equiv I(E_{44'} + \mathcal{E}_{4'2''}),$$

¹⁰ Compare Bolza, *Vorlesungen*, pp. 405-07.

and define

$$S(\tau_{2''}) \equiv W(x_{2''}, y_{2''}).$$

It can be shown that

$$S'(\tau_{2''}) = X'_2 F_{x'}(x, \phi') + Y'_2 F_{y'}(x, \phi')$$

at $2''$, which is positive near 2 . It is easily seen that

$$I(\mathcal{E}_{42'}) - I(E_{42}) = S(\tau_{2'}) - S(\tau_2)$$

if the point 4 precedes 2 along E_{12} . Therefore the inequality (13) holds in this case if \mathcal{F}_2 is sufficiently restricted. In the same way we can show that this inequality holds if C is interior to R for $\tau \leq \tau_2$.

If $D(x', X'_2) = 0$ at 2 and C lies beyond the normal to E_{12} at 2 in such a way that the point 4 lies beyond 2 along E_{56} , then the inequality (11) becomes

$$(14) \quad I(V_{1'2'}) > I(\mathcal{E}_{1'3} + E_{32}) + I(E_{24} + \mathcal{E}_{42'}).$$

An argument similar to the one just given can be used to show that

$$(15) \quad I(E_{24} + \mathcal{E}_{42'}) > 0,$$

in this case if the field \mathcal{F}_2 is sufficiently restricted. From the inequalities (11) and (13) or from (14) and (15), it follows that

$$(16) \quad I(V_{1'2'}) > I(\mathcal{E}_{1'3} + E_{32}).$$

Steps perfectly analogous to those just completed, but making use of the field \mathcal{F}_1 defined by the family of extremals (\mathcal{E}) preceding their points of tangency with E_{56} , can be taken to prove that the inequality (12) is true if \mathcal{F}_1 is sufficiently restricted and the point 3 follows 1 along E_{56} . If $D(x', X'_1) = 0$ at 1 and the curve D precedes the normal to E_{12} at 1 in such a way that 3 precedes 1 on E_{56} , then the inequality (16) becomes

$$(17) \quad I(V_{1'2'}) > I(\mathcal{E}_{1'3} + E_{31} + E_{12}),$$

and it can be shown as in the analogous case at 2 that

$$(18) \quad I(\mathcal{E}_{1'3} + E_{31}) > 0.$$

From the inequalities (12) and (16) or from (17) and (18), it follows that

$$I(V_{1'2'}) > I(E_{12}).$$

This completes the proof of Theorem 2 under the assumption that there is some point 8 on E_{12} at which $F(x, x') \neq 0$. If E_{12} is a solution of the differential equation $F(x, x') = 0$, that is if F is identically zero along E_{12} , one may replace the field \mathcal{F}_8 used in the above proof by one determined in a neighborhood of 8 by the one parameter family of extremals through the point 8, for example.

It should be pointed out that if $F(x, x') = 0$ and not both $F_{x'}$ and $F_{y'}$ are zero at 1 or 2, then Theorem 2 may be vacuous as follows from the geometric interpretation of the conditions of the Theorem already given. For example, if this is true at the point 2, then the curve C is not tangent to E_{12} and $\tilde{T}(x, x', X'_2)$ is positive or is negative for all directions $(X'_2, Y'_2) \neq (kx', ky')$ ($k \neq 0$). In the latter case there is no direction (X'_2, Y'_2) for which the conditions of the Theorem are satisfied.

The next four theorems have to do with the case where $T = 0$ along E_{12} , that is when E_{12} is an arc of an extremal. Therefore, each theorem will involve the Jacobi condition for the arc E_{12} . By means of the methods used by the author¹¹ and those of Bliss¹² for two variable end-points, one is able to derive the Jacobi condition for each of the cases. These proofs do not require that the enveloping curve have a regressive branch at its point of contact with the minimizing curve. We shall state these conditions strengthened in the usual way.

(I) E_{12} contains no pair of conjugate points.

If the curve D is transversal to E_{12} at the point 2, then

(II) E_{12} does not contain the critical point of the curve D .

If the curve C is transversal to E_{12} at the point 2, then

(III) E_{12} does not contain the critical point of the curve C .

If the curves D and C are transversal to E_{12} at the points 1 and 2 respectively, then

(IV) The critical point d of the curve D and the critical point c of the curve C lie in the order 12cd or some cyclic order with all points distinct.

¹¹ Mancill, "The Jacobi condition for unilateral variations," *Duke Journal of Mathematics*, vol. 6 (1940), pp. 341-44.

¹² *Loc. cit.*, p. 74.

The next theorem is analogous to Theorem 2 for the case where E_{12} is an arc of an extremal.

THEOREM 3. *Suppose that E_{12} is an extremal arc with the properties:*

- 1) *condition (I) is satisfied by E_{12} ;*
- 2) *if $D(x', X'_1) \neq 0$ at 1, then $\bar{T}(x, x', X'_1) < 0$ there; or if $D(x', X'_1) = 0$ at 1, then*

$$F(x, x') > 0 \text{ at 1 if } D \text{ precedes the normal to } E_{12} \text{ at 1,}$$

$$F(x, x') < 0 \text{ at 1 if } D \text{ lies beyond the normal to } E_{12} \text{ at 1;}$$
- 3) *if $D(x', X'_2) \neq 0$ at 2, then $\bar{T}(x, x', X'_2) > 0$ there; or if $D(x', X'_2) = 0$ at 2, then*

$$F(x, x') < 0 \text{ at 2 if } C \text{ precedes the normal to } E_{12} \text{ at 2,}$$

$$F(x, x') > 0 \text{ at 2 if } C \text{ lies beyond the normal to } E_{12} \text{ at 2.}$$

Then the arc E_{12} minimizes the integral I in the class of all admissible curves sufficiently near E_{12} joining the curves D and C .

There exists a one parameter family of extremals

$$x = \phi(t, a), \quad y = \psi(t, a),$$

through a point 0 preceding the point 1 on E_{12} , containing E_{12} for $a = a_0$, and which forms a field \mathcal{F} determined by the parameter values

$$a_0 \leq a \leq a_0 + e, \quad t_1 - e \leq t \leq t_2 + e,$$

for e sufficiently small. This field lies interior to the region R of admissible curves. If we assume that the curve D is inferior to R for $\tau \geq \tau_1$, and that C is interior to R for $\tau \geq \tau_2$, then it follows from the hypotheses of the Theorem that

$$X'_1 F_{x'}(x, x') + Y'_1 F_{y'}(x, x') < 0$$

and

$$X'_2 F_{x'}(x, x') + Y'_2 F_{y'}(x, x') > 0$$

at 1 and 2 respectively. Let $V_{1'2'}$ be any admissible comparison curve in \mathcal{F} joining the points 1' and 2' respectively. Then we have

$$\begin{aligned}
 (19) \quad I(V_{1'2'}) - I(E_{12}) &= I(V_{1'2'}) - I^*(E_{12}) \\
 &= I(V_{1'2'}) - I^*(V_{1'2'}) + I^*(C_{22'}) - I^*(D_{11'})
 \end{aligned}$$

because of the invariance of the integral I^* . But the right member of (19) is positive since our problem is regular and

$$I^*(D_{11'}) < 0$$

and

$$I^*(C_{22'}) > 0,$$

if the field \mathcal{F} is sufficiently restricted. Obviously the same conclusion can be reached if the positive direction along either D or C is reversed. Therefore, Theorem 3 is proved.

THEOREM 4. Suppose that E_{12} is an extremal arc with the properties:

- 1) $\tilde{T}(x, x', X'_2) = 0$ and $F(x, x') \neq 0$ at the point 2;
- 2) condition (II) is satisfied by E_{12} ;
- 3) same as in Theorem 3.

Then the arc E_{12} minimizes the integral I in the class of all admissible curves sufficiently near E_{12} joining the curves D and C .

The one parameter family of extremals cut by D transversally forms a field in R in which we may apply the same proof as in Theorem 3, since C is not transversal to E_{12} at the point 2. In the same way we can prove the next theorem.

THEOREM 5. Suppose that E_{12} is an extremal arc with the properties:

- 1) $\tilde{T}(x, x', X'_2) = 0$ and $F(x, x') \neq 0$ at the point 2;
- 2) same as in Theorem 3;
- 3) condition (III) is satisfied by E_{12} .

Then the arc E_{12} minimizes the integral I in the class of all admissible curves sufficiently near E_{12} joining the curves D and C .

In the next two theorems we shall treat the case analogous to the problem of free variations with both end-points variable,¹³ that is where E_{12} is an extremal and both D and C are transversal to E_{12} . The one parameter family of extremals

$$(20) \quad x = \phi(t, a), \quad y = \psi(t, a),$$

cut by D transversally contains E_{12} for $a = a_0$. If this family forms a field about E_{12} , we shall denote by C' the transversal of this field through the point 2, which in this case will be tangent to C at 2.

THEOREM 6. *Suppose that E_{12} is an extremal arc with the properties:*

1) $\tilde{T}(x, x', X'_1) = 0$ and $F(x, x') \neq 0$ at the point 1;

2) condition (II) is satisfied by E_{12} ;

3) $\tilde{T}(x, x', X'_2) = 0$ and $F(x, x') \neq 0$ at the point 2;

4) $F(x, x') > 0$ at 2 and C' coincides with or lies on the same side of C as does D near 2 in R ; or $F(x, x') < 0$ at 2 and C' coincides with C or lies on the side of C opposite to D near 2 in R .

Then E_{12} minimizes the integral I in the class of all admissible curves joining the curves D and C and sufficiently near E_{12} .

Under the hypotheses of the theorem the family of extremals (20) forms a field about E_{12} . Also there is a unique curve C' through 2 which cuts each of the extremals of the field transversally. Let $V_{1'2'}$ denote any admissible comparison curve in the field intersecting the curve D and C in $1'$ and $2'$ respectively. Suppose that the first part of hypothesis 4) of the theorem is satisfied. Then the usual proof shows that

$$(21) \quad I(V_{1'2'}) > I(\mathcal{E}_{52'}),$$

where $\mathcal{E}_{52'}$ is the particular member of the family (20) through the point 5 on D and the point $2'$ on C . Let the extremal $\mathcal{E}_{52'}$ meet C' in the point 3, and we have

$$I(\mathcal{E}_{52'}) = I(\mathcal{E}_{53} + \mathcal{E}_{32'}) = I(E_{12} + \mathcal{E}_{32'}),$$

since D and C' are both transversals of the field. Since $F(x, x') > 0$ at 2,

¹³ For the problem in non-parametric form see, Bliss, *loc. cit.*, pp. 75-80.

it will remain positive along each member of the family (20) in a sufficiently small neighborhood of 2. Therefore, if the field is sufficiently restricted, we have

$$I(\mathcal{E}_{32'}) \geq 0.$$

Consequently the theorem is proved for the case when C' precedes C or coincides with C along the extremals (20) and $F(x, x') > 0$ at 2.

Now suppose that the second part of hypothesis 4) of the theorem is satisfied. In this case $F(x, x') < 0$ at 2 and

$$I(\mathcal{E}_{2'3}) < 0$$

if the field is sufficiently restricted. Therefore, it follows from the inequality (21) that

$$I(V_{1'2'}) > I(\mathcal{E}_{32'} + \mathcal{E}_{2'3}) = I(E_{12}),$$

since D and C are transversals of the field.

THEOREM 7. *Suppose that E_{12} is an extremal arc with the properties:*

- 1) $\bar{T}(x, x', X'_1) = 0$ and $F(x, x') \neq 0$ at the point 1;
- 2) $\bar{T}(x, x', X'_2) = 0$ and $F(x, x') \neq 0$ at the point 2;
- 3) condition (IV) is satisfied along E_{12} ;

Then E_{12} minimizes the integral I in the class of all admissible curves joining the curves D and C and sufficiently near E_{12} .

Again the family of extremals (20) forms a field about E_{12} . It has been shown in general that the position of the critical point c of the curve C depends only upon the curvature of C at its intersection 2 with E_{12} . The radius of curvature r of C at 2 is thus a known function of the parameter t of the critical point c . This function has a non-zero derivative dr/dt whose sign is that of

$$F(x, x') \sin(\theta_E - \theta_C)/F_1(x, x')$$

at 2, where θ_E and θ_C are the angles that E_{12} and C respectively make with the x -axis.¹⁴ Now suppose we consider the case where the order of points in condition (IV) is 12cd. We shall break the proof into two cases:

¹⁴ Bliss, "The second variation of a definite integral when one end-point is variable," *Transactions of the American Mathematical Society*, vol. 3 (1902), p. 139. There seems to be an error in sign in the formula for dr/dt on this page.

1) $F(x, x') > 0$ at the point 2:

If C is directed toward the left of E_{12} , that is toward the interior of R at 2, then the derivative $dr/dt < 0$ and $r_c > r_{c'}$, since c precedes d and d is the critical point of C' . But in this case r_c and $r_{c'}$ are negative, and therefore C' lies to the left of C , that is on the same side of C as D . If C is directed toward the right of E_{12} , then the derivative $dr/dt > 0$ and $r_c < r_{c'}$ since c precedes d . But in this case r_c and $r_{c'}$ are positive, and therefore C' lies on the same side of C as D . Thus, Theorem 6 applies in this case.

2) $F(x, x') < 0$ at the point 2:

If C is directed toward the left of E_{12} , the derivative $dr/dt > 0$ and $r_c < r_{c'}$ since c precedes d . But in this case r_c and $r_{c'}$ are positive and C' lies on the side of C opposite to D . If C is directed toward the right of E_{12} , then $dr/dt < 0$ and $r_c > r_{c'}$ since c precedes d . But in this case r_c and $r_{c'}$ are negative and C' lies on the side of C opposite to D . Therefore, Theorem 6 applies again.

A similar proof can be made for each of the cases determined by the order of the points 1, 2, c , and d . If the points c and d coincide, then E_{12} may or may not minimize the integral I , but Theorem 6 furnishes a set of sufficient conditions in this case.

The treatment of the case where E_{12} is an extremal and the curves D and C are transversal to E_{12} just completed is the same for free variations. It is the opinion of the author that the treatment given here is simpler than that given by Bliss to which we have already referred.

UNIVERSITY OF ALABAMA.

ON SOME TRIANGULAR SUMMABILITY METHODS.*

By JOSHUA BARLAZ.

1. Let there be given a sequence-to-function transformation

$$(1.1) \quad \Psi(x) = \sum_{\nu=0}^{\infty} s_{\nu} \psi_{\nu}(x),$$

where $\{\psi_{\nu}(x)\}$ is a sequence of functions defined on a point set \mathcal{E} having a limit point ξ not belonging to the set; then the generalized limit of the sequence $\{s_{\nu}\}$ is defined by

$$\lim_{x \rightarrow \xi} \Psi(x) = s = \text{gen lim } s_{\nu}$$

if the limit on the left exists. In a recent paper,¹ Otto Szász considered the question of selecting a sequence of points $\{x_n\}$ belonging to the point set \mathcal{E} , $x_n \rightarrow \xi$ as $n \rightarrow \infty$, and associating with it the sequence-to-sequence triangular transformation

$$(1.2) \quad B_n(x_n) = \sum_{\nu=0}^n s_{\nu} \psi_{\nu}(x_n), \quad n = 0, 1, 2, \dots,$$

where the generalized limit of the sequence $\{s_{\nu}\}$ is now defined by

$$\lim_{n \rightarrow \infty} B_n(x_n) = s = \text{gen lim } s_{\nu}.$$

It is to be noted that $B_n(x)$ is defined by taking the first $n + 1$ terms of the series defining the function $\Psi(x)$; the summability method is then constructed with the sequence $\{x_n\}$.

As set forth in the aforementioned paper of Szász,² the regularity of either method (1.1) or (1.2) does not imply the regularity of the other method. On the other hand, if the first method is boundedness preserving, then the associated triangular methods are also boundedness preserving for all sequences $\{x_n\}$, $x_n \rightarrow \xi$. The conditions of regularity for the transformation (1.2) are

$$(1.3) \quad \lim_{n \rightarrow \infty} \psi_{\nu}(x_n) = 0, \quad \nu = 0, 1, 2, \dots,$$

* Received February 27, 1946. Presented to the American Mathematical Society, September 15, 1945. The author wishes to thank Professor Otto Szász for his interest in the preparation of this paper.

¹ The superscripts refer to the notes and references at the end of the paper.

$$(1.4) \quad \sum_{\nu=0}^n |\psi_{\nu}(x_n)| < K, \quad K \text{ independent of } n,$$

$$(1.5) \quad \lim_{n \rightarrow \infty} \sum_{\nu=0}^n \psi_{\nu}(x_n) = 1.$$

The purpose of this paper is to consider the triangular matrix transformation which results when $\psi_{\nu}(x)$ is taken equal to $e^{-x}x^{\nu}/\nu!$; that is, when the generating process is applied to Borel's exponential method. The generalized limit of the sequence $\{s_{\nu}\}$ is now defined by

$$(1.6) \quad \text{gen lim } s_{\nu} = \lim_{n \rightarrow \infty} e^{-x_n} \sum_{\nu=0}^n (s_{\nu} x_n^{\nu}/\nu!), \quad x_n \rightarrow \infty.$$

2. The first problem considered for this summability method is the determination of necessary and sufficient conditions on the sequence of points $\{x_n\}$, $x_n \rightarrow \infty$, so that the regularity conditions (1.3), (1.4), and (1.5) are satisfied. It is easily seen that the conditions (1.3) and (1.4) are here fulfilled with every sequence $\{x_n\}$, so long as $x_n \rightarrow \infty$. Therefore, in order to insure regularity, it remains to choose the sequence $\{x_n\}$ so that

$$(2.1) \quad g_n(x_n) \equiv g_n = e^{-x_n} \sum_{\nu=0}^n x_n^{\nu}/\nu! \rightarrow 1 \text{ as } n \rightarrow \infty.$$

In determining the sequences $\{x_n\}$ satisfying (2.1), we shall use the following limit due to Gauss:³

$$(A) \quad \lim_{x \rightarrow \infty} e^{-x} \sum_{\nu < x + \omega \sqrt{x}} x^{\nu}/\nu! = 1/\sqrt{2\pi} \int_{-\infty}^{\omega} e^{-t^2/2} dt \equiv p(\omega).$$

The function $p(\omega)$ has the following properties:

1. $p(-\infty) = \lim_{\omega \rightarrow -\infty} p(\omega) = 0,$
2. $p(0) = 1/2,$
3. $p(\infty) = \lim_{\omega \rightarrow \infty} p(\omega) = 1,$
4. $p(\omega)$ increases continuously as ω increases.

THEOREM 2.1. *A necessary and sufficient condition that*

$$(2.2) \quad g_n(x_n) \rightarrow \lambda \text{ as } n \rightarrow \infty, \quad 0 \leq \lambda \leq 1$$

is that

$$(2.3) \quad \lim_{n \rightarrow \infty} (x_n - n)/\sqrt{n} = -\rho, \quad x_n \rightarrow \infty$$

where $p(\rho) = \lambda$, i. e., $\rho = p_{-1}(\lambda)$.

Proof. (a) Sufficiency: We consider first the case when ρ is finite. It follows from (2.3) that for every positive ϵ , however small, there exists an $N_0 = N_0(\epsilon)$ such that

$$n - (\rho + \epsilon)\sqrt{n} \leq x_n \leq n - (\rho - \epsilon)\sqrt{n}, \quad n > N_0.$$

Transforming the above inequalities, we have

$$(\sqrt{n} - (\rho + \epsilon)/2)^2 - (\rho + \epsilon)^2/4 \leq x_n \leq (\sqrt{n} - (\rho - \epsilon)/2)^2 - (\rho - \epsilon)^2/4.$$

Choosing n so large that $\sqrt{n} > (\rho + \epsilon)/2$, we obtain

$$\begin{aligned} \sqrt{x_n + (\rho - \epsilon)^2/4} + (\rho - \epsilon)/2 &\leq \sqrt{n} \leq \sqrt{x_n + (\rho + \epsilon)^2/4} + (\rho + \epsilon)/2 \\ x_n + (\rho - \epsilon)\sqrt{x_n + (\rho - \epsilon)^2/4} + (\rho - \epsilon)^2/2 \\ &\leq n \leq x_n + (\rho + \epsilon)\sqrt{x_n + (\rho + \epsilon)^2/4} + (\rho + \epsilon)^2/2. \end{aligned}$$

Inasmuch as $x_n \rightarrow \infty$, we have for n large enough

$$x_n + (\rho - 2\epsilon)\sqrt{x_n} \leq n \leq x_n + (\rho + 2\epsilon)\sqrt{x_n}.$$

From this it immediately follows that

$$(2.4) \quad e^{-x_n} \sum_{\nu \leq x_n + (\rho - 2\epsilon)\sqrt{x_n}} x_n^\nu / \nu! \leq g_n \leq e^{-x_n} \sum_{\nu \leq x_n + (\rho + 2\epsilon)\sqrt{x_n}} x_n^\nu / \nu!;$$

letting n become infinite, we get from (A)

$$p(\rho - 2\epsilon) \leq \liminf g_n \leq \limsup g_n \leq p(\rho + 2\epsilon).$$

Inasmuch as ϵ may be chosen arbitrarily small, and $p(\omega)$ is continuous, we have

$$\lim g_n(x_n) = p(\rho) = \lambda.$$

If ρ is positive infinity, the condition (2.3) leads to the inequality

$$n \geq x_n + K\sqrt{x_n},$$

K arbitrarily large, n large enough depending on K . This implies that

$$p(K) \leq \liminf g_n \leq \limsup g_n \leq 1,$$

whence it follows that

$$\lim g_n = p(\infty) = 1.$$

If ρ is negative infinity, we obtain similarly

$$n \leq x_n - K\sqrt{x_n},$$

and consequently $0 \leq \liminf g_n \leq \overline{\lim} g_n \leq p(-K)$ which yields $\lim g_n = p(-\infty) = 0$. This completes the proof of the sufficiency.

(b) Necessity: The necessity of (2.3) for (2.2) follows easily from (a), the proof of the sufficiency; for, by (a), the sequence $\{(x_n - n)/\sqrt{n}\}$ could not have more than one limit point, finite or infinite, in order that g_n have a limit. Again by (a), the one limit point must be $-\rho$.

THEOREM 2.2. *A necessary and sufficient condition that the summability method defined by (1.6) be regular is that*

$$(2.5) \quad \lim_{n \rightarrow \infty} (x_n - n)/\sqrt{n} = -\infty, \quad x_n \rightarrow \infty \text{ as } n \rightarrow \infty.$$

This theorem follows immediately from the preceding one and the fact that (1.3) and (1.4) are satisfied when $x_n \rightarrow \infty$.

It is easy to see that we can derive from Theorem 2.1 the following result: Consider the summability method defined by the relation

$$(2.6) \quad \text{gen lim } s_\nu = \lim_{n \rightarrow \infty} e^{-x_n/\lambda} \sum_{\nu=0}^n s_\nu x_n^\nu / \nu!, \quad x_n \rightarrow \infty, \quad 0 < \lambda \leq 1.$$

A necessary and sufficient condition for the regularity of the method is (2.3). Theorem 2.2 is the special case $\lambda = 1$, $\rho = \infty$.

The transform (1.6) clearly is always boundedness preserving. If $\lim (x_n - n)n^{-1/2}$ exists, the transform is convergence preserving. Furthermore, if $\lim (x_n - n)n^{-1/2} = \infty$, the method sums every bounded sequence to the value zero.

3. The first application that we shall discuss for the transform (1.6) is to analytic functions of a complex variable defined by a power series with a finite radius of convergence. It is the purpose of this section to show that, with certain conditions on the sequence of points $\{x_n\}$, the summability method defined by (1.6) yields the analytic continuation of the function within the Borel polygon.^{4,5}

Let there be given a function defined by a power series in a positive circle of convergence:

$$(3.1) \quad f(z) = \sum_{\nu=0}^{\infty} a_\nu z^\nu, \quad |z| < r.$$

Let

$$(3.2) \quad s_n(z) = \sum_{\nu=0}^n a_\nu z^\nu, \quad n = 0, 1, 2, \dots,$$

$$(3.3) \quad F(z, x_n) \equiv F_n(z) = e^{-x_n} \sum_{\nu=0}^n s_\nu(z) x_n^\nu / \nu!.$$

Then it is clear that (3.3) gives the application of (1.6) to the sequence of partial sums defined in (3.2). When $\{x_n\}$ satisfies the condition (2.5), Theorem 2.2 implies that $F_n(z) \rightarrow f(z)$ as $n \rightarrow \infty$ for $|z| < r$. As has been anticipated, the object is to determine additional conditions on $\{x_n\}$ so that $F_n(z) \rightarrow f(z)$ as $n \rightarrow \infty$ in a larger region for z , namely the Borel polygon for $f(z)$. We first proceed to set up several formulae relating $F_n(z)$ to $f(z)$ more directly than is done by the defining formula (3.3).

Having started with the function $f(z)$, defined by (3.1), let Γ be any closed Jordan curve containing the origin, such that $f(z)$ is regular interior to and on Γ . Then from the formula

$$(3.4) \quad a_m = 1/2\pi i \int_{\Gamma} \frac{f(z)}{z^{m+1}} dz, \quad m = 0, 1, 2, \dots,$$

it follows that

$$s_\nu(z) = \sum_{m=0}^{\nu} a_m z^m = \sum_{m=0}^{\nu} z^m / 2\pi i \int_{\Gamma} \frac{f(\xi)}{\xi^{m+1}} d\xi,$$

$$s_\nu(z) = 1/2\pi i \int_{\Gamma} \frac{f(\xi)}{\xi} \left\{ \frac{1 - z^{\nu+1}/\xi^{\nu+1}}{1 - z/\xi} \right\} d\xi, \quad z \text{ not on } \Gamma.$$

Assume now that z is interior to Γ . We obtain

$$s_\nu(z) = 1/2\pi i \int_{\Gamma} \frac{f(\xi)}{\xi - z} d\xi - 1/2\pi i \int_{\Gamma} \frac{f(\xi) \cdot z^{\nu+1}}{(\xi - z)\xi^{\nu+1}} d\xi,$$

$$(3.5) \quad s_\nu(z) = f(z) - 1/2\pi i \int_{\Gamma} \frac{f(\xi) \cdot z^{\nu+1}}{(\xi - z)\xi^{\nu+1}} d\xi.$$

Substituting now the formula (3.5) for $s_\nu(z)$ into (3.3), we get

$$F_n(z) = e^{-x_n} \sum_{\nu=0}^n f(z) x_n^\nu / \nu! - e^{-x_n} / 2\pi i \sum_{\nu=0}^n x_n^\nu / \nu! \int_{\Gamma} \frac{f(\xi) z^{\nu+1}}{(\xi - z)\xi^{\nu+1}} d\xi,$$

$$(3.6) \quad F_n(z) = e^{-x_n} \sum_{\nu=0}^n f(z) x_n^\nu / \nu! - e^{-x_n} / 2\pi i \int_{\Gamma} \frac{f(\xi) z}{(\xi - z)\xi} \left\{ \sum_{\nu=0}^n x_n^\nu z^\nu / \xi^{\nu+1} \right\} d\xi,$$

or

$$(3.6') \quad F_n(z) \equiv f(z) g_n(x_n) - I(x_n, z),$$

where

$$(3.7) \quad I(x_n, z) \equiv I_n = e^{-x_n} / 2\pi i \int_{\Gamma} \frac{f(\xi) z}{(\xi - z)\xi} \left\{ \sum_{\nu=0}^n x_n^\nu z^\nu / \xi^{\nu+1} \right\} d\xi,$$

and $g_n(x_n)$ is given in (2.1). Thus the problem has been reduced to the determination of conditions on $\{x_n\}$ so that $g_n \rightarrow 1$, $I_n \rightarrow 0$ as $n \rightarrow \infty$. Clearly $F_n(z) \rightarrow f(z)$ under these conditions.

We now proceed to simplify the integral I_n by means of a transformation of the kernel $\sum_{\nu=0}^n x_n^\nu z^\nu / \zeta^\nu \nu!$ into an integral. As has become standard in such cases, the starting point for the conversion is the familiar formula

$$m! = \int_0^\infty t^m e^{-t} dt. \quad \text{We have}$$

$$(3.8) \quad K(\lambda) = 1/n! \sum_{\nu=0}^n \frac{n! \lambda^\nu}{\nu! (n-\nu)!} \int_0^\infty t^{n-\nu} e^{-t} dt = \sum_{\nu=0}^n \lambda^\nu / \nu!,$$

$$K(\lambda) = 1/n! \int_0^\infty (\lambda + t)^n e^{-t} dt, \quad n \geq 0.$$

Write $z/\zeta = \omega = \alpha + i\beta$, α, β real. Then

$$\begin{aligned} K_n &\equiv K(x_n z / \zeta) = 1/n! \int_0^\infty (t + (x_n z / \zeta))^n e^{-t} dt \\ &= 1/n! \int_0^\infty (t + \alpha x_n + i\beta x_n)^n e^{-t} dt; \\ n! K_n &= e^{\alpha x_n} \int_{\alpha x_n}^\infty (t + ix_n \beta)^n e^{-t} dt \\ &= e^{x_n(\alpha + i\beta)} \int_{\alpha x_n}^\infty (t + ix_n \beta)^n e^{-(t + ix_n \beta)} dt. \end{aligned}$$

Considering $t + ix_n \beta$ as a complex variable s , we have

$$e^{-\omega x_n n} K_n = \int_C e^{-s} s^n ds$$

where C is the straight line path $x_n \alpha + ix_n \beta \rightarrow +\infty + ix_n \beta$. Let $C(\rho)$ designate the straight line path $x_n \alpha + ix_n \beta \rightarrow \rho + ix_n \beta$. Then

$$(3.9) \quad e^{-\omega x_n n} K_n = \lim_{\rho \rightarrow \infty} \int_{C(\rho)} e^{-s} s^n ds.$$

From the fact that $e^{-s} s^n$ is a regular function of s in the whole complex plane, it follows that

$$(3.10) \quad \int_{C(\rho)} e^{-s} s^n ds = \left(\int_{C_1} + \int_{C_2} + \int_{C_3} \right) s^n e^{-s} ds \equiv T_1 + T_2 + T_3$$

where C_1, C_2, C_3 are the straight line paths $x_n \omega \rightarrow 0, 0 \rightarrow \rho, \rho \rightarrow \rho + ix_n \beta$ respectively. It is easily seen that

$$(3.11) \quad |T_3| \leq |x_n \beta| e^{-\rho} |\rho + ix_n \beta|^n = o(1) \text{ as } \rho \rightarrow \infty,$$

$$(3.12) \quad T_2 = \int_0^\rho e^{-s} s^n ds \rightarrow n! \text{ as } \rho \rightarrow \infty.$$

We now have from (3.9), (3.10), (3.11), (3.12)

$$(3.13) \quad K_n = \sum_{\nu=0}^n x_n^\nu z^\nu / \xi^\nu \nu! = e^{\omega x_n} + (e^{\omega x_n} / n!) \int_{C_1} s^n e^{-s} ds.$$

Combining (3.7) and (3.13), we obtain

$$(3.14) \quad I_n = 1/2\pi i \int_{\Gamma} \frac{f(\xi)z}{(\xi-z)\xi} e^{-x_n(1-z/\xi)} d\xi + R_n(z),$$

where

$$(3.15) \quad R_n(z) = 1/2\pi i \int_{\Gamma} \frac{f(\xi)z}{(\xi-z)\xi} e^{-x_n(1-z/\xi)} \{1/n! \int_{C_1} s^n e^{-s} ds\} d\xi.$$

THEOREM 3.1. *If $x_n = o(n)$ as $n \rightarrow \infty$, then $R_n(z) \rightarrow 0$ as $n \rightarrow \infty$, for all z (not on Γ).*

Proof. From the hypothesis, it follows that for any positive ϵ , however small, $x_n \leq \epsilon n$ for n large enough. Using Stirling's formula: $n! \approx n^n e^{-n} (2\pi)^{1/2}$, we get

$$\begin{aligned} |e^{-x_n(1-\omega)} / n! \int_{C_1} s^n e^{-s} ds| &\leq e^{-x_n(1-a)} e^{|a|x_n} |x_n z / \xi|^{n+1} / n! \\ &\leq K \frac{e^{Ax_n} B^{n+1} n^{n+1} \epsilon^{n+1}}{n^n e^{-n} (2\pi n)^{1/2}}, \quad n > n_0. \end{aligned}$$

Recalling that Γ is a closed Jordan curve containing the origin, we see that K, A, B are absolute constants independent of ξ, ζ on Γ . Hence, if ϵ is chosen small enough, it follows that the right hand member of the above inequality tends to zero uniformly as to ξ, ζ on Γ , as $n \rightarrow \infty$. This is sufficient to prove the theorem.^a

THEOREM 3.2. *If $x_n = o(n)$, $x_n \rightarrow \infty$, as $n \rightarrow \infty$, then $F_n(z) \rightarrow f(z)$ as $n \rightarrow \infty$, for all z in the Borel polygon of the function $f(z)$.*

Proof. From the condition $x_n = o(n)$, Theorem 2.1 implies that $g_n \rightarrow 1$ as $n \rightarrow \infty$. If z satisfies the inequality $\mathcal{R}(z/\xi) < 1$ for all ξ on Γ , clearly

$$(3.16) \quad 1/2\pi i \int_{\Gamma} \frac{f(\xi)z}{(\xi-z)\xi} e^{-x_n(1-z/\xi)} d\xi \rightarrow 0, \quad n \rightarrow \infty.$$

It is well known that the condition $\mathcal{R}(z/\xi) < 1$ for all ξ on Γ implies that,

with the proper choice of the curves Γ , (3.16) can be proved for all z in the polygon.⁷ Applying the formulae (3.6'), (3.14), and Theorem 3.1, the proof is rendered complete.

Theorem 3.2 gives a condition under which the triangular summability method defined by (1.6) sums a power series within the Borel polygon of the function represented by this series.⁸ However, we can also use the formulae derived on the preceding pages to show that if $x_n = o(n)$, then the transform (1.6) is as powerful as Borel's exponential method of summation when applied to a power series with a non-zero radius of convergence.

To demonstrate the result just mentioned, let us return to formula (3.5). We observe that if z is exterior to Γ , then (3.5) becomes

$$(3.17) \quad s_\nu(z) = -1/2\pi i \int_{\Gamma} f(\xi) z^{\nu+1}/(\xi-z)\xi^{\nu+1} d\xi.$$

Using (3.3), (3.5), (3.6'), (3.14), (3.17), we can now write for all z not on Γ

$$(3.18) \quad F(z, x_n) = \delta(z, \Gamma)f(z)g_n(x_n) - 1/2\pi i \int_{\Gamma} \frac{f(\xi)z}{(\xi-z)\xi} e^{-x_n(1-z/\xi)} d\xi - R_n(z)$$

where Γ is any closed Jordan curve containing the origin, such that $f(z)$ is regular interior to and on Γ , and

$$(3.19) \quad \begin{aligned} \delta(z, \Gamma) &= 1, \text{ if } z \text{ is interior to } \Gamma \\ &= 0, \text{ if } z \text{ is exterior to } \Gamma. \end{aligned}$$

The equation giving the application of Borel's exponential method to the series in (3.1) is

$$(3.20) \quad B(z, x) = \delta(z, \Gamma)f(z) - 1/2\pi i \int_{\Gamma} \frac{f(\xi)z}{(\xi-z)\xi} e^{-x_n(1-z/\xi)} d\xi, \quad x \rightarrow \infty.$$

Combining (3.18) and (3.20), we have at once

$$B(z, x_n) - F(z, x_n) = \delta(z, \Gamma)f(z)[1 - g_n] + R_n(z).$$

Using Theorems 2.1 and 3.1, we have

THEOREM 3.3. *If $x_n = o(n)$, then the transform (1.6) is as powerful as Borel's method in summing power series with a positive radius of convergence.*

4. In this section we consider the application of the transform (1.6) to Fourier series. It is immediately seen from Theorem 3.3 that if $x_n = o(n)$, then the triangular method will sum a Fourier series whenever Borel's method does. However, the proof of Theorem 3.2 and the footnote referred to therein suggest the possibility of less restrictive conditions on the sequence $\{x_n\}$ in the case of Fourier series.

Let $f(\theta)$ be a Lebesgue integrable function of period 2π . Following the usual notation for Fourier series, we have

$$(4.1) \quad f(\theta) \sim a_0/2 + \sum_1^{\infty} (a_n \cos n\theta + b_n \sin n\theta) \equiv \sum_0^{\infty} A_n(\theta).$$

$$(4.2) \quad A_n(\theta) \rightarrow 0 \text{ as } n \rightarrow \infty, \text{ uniformly as to } \theta.$$

$$(4.3) \quad \pi a_n = \int_0^{2\pi} f(t) \cos nt dt, \quad \pi b_n = \int_0^{2\pi} f(t) \sin nt dt, \quad n \geq 0.$$

Writing

$$s_n(\theta) = \sum_{\nu=0}^n A_{\nu}(\theta), \quad n \geq 0, \quad s_n^*(\theta) = s_n(\theta) - \frac{1}{2}A_n(\theta), \quad n > 0, \quad s_0^* = 0,$$

it follows that

$$(4.4) \quad s_{\nu}^*(\theta) = 1/\pi \int_0^{\pi} \psi(\theta, t) \cot(t/2) \sin \nu t dt, \text{ where}$$

$$(4.5) \quad \psi(\theta, t) = \frac{1}{2}\{f(\theta+t) + f(\theta-t)\}.$$

From (4.2) and the fact that the method (1.6) is linear and boundedness preserving for all sequences $\{x_n\}$, therefore preserving convergence to zero as well as uniform convergence to zero, it follows that we may use the sequence $\{s_n^*(\theta)\}$ instead of $\{s_n(\theta)\}$ in the formal application of (1.6).

Write

$$(4.6) \quad P_n \equiv P(\theta, x_n) = e^{-x_n} \sum_{\nu=0}^n s_{\nu}^*(\theta) x_n^{\nu} / \nu!, \quad n = 0, 1, 2, \dots$$

Then from (4.4)

$$(4.7) \quad P_n = e^{-x_n} / \pi \int_0^{\pi} \psi(\theta, t) \cot(t/2) \left\{ \sum_{\nu=0}^n x_n^{\nu} \sin \nu t / \nu! \right\} dt.$$

Employing (3.8) and (3.13), we obtain

$$\begin{aligned}\sum_{\nu=0}^n x_n^\nu \sin \nu t / \nu! &= \mathfrak{A} \left\{ \sum_{\nu=0}^n x_n^\nu e^{i\nu t} / \nu! \right\} = \mathfrak{A} \{ K(x_n e^{it}) \} \\ &= \mathfrak{A} \{ \exp(x_n e^{it}) \} + \mathfrak{A} \{ \exp(x_n e^{it}) \int_{C_1} s^n e^{-s} / n! ds \},\end{aligned}$$

where C_1 is the straight line path $x_n e^{it} \rightarrow 0$. Therefore

$$\begin{aligned}(4.8) \quad P_n &= 1/\pi \int_0^\pi \psi(\theta, t) \cot(t/2) e^{-x_n(1-\cos t)} \sin(x_n \sin t) dt + \Omega_n \\ &\equiv B(\theta, x_n) + \Omega_n\end{aligned}$$

where

$$(4.9) \quad \Omega_n = 1/\pi \int_0^\pi \psi(\theta, t) \cot(t/2) \mathfrak{A} \{ \exp(-x_n + x_n e^{it}) \int_{C_1} s^n e^{-s} / n! ds \} dt,$$

and $B(\theta, x)$ is the same as the familiar expression which is obtained when Borel's exponential mean is applied to Fourier series.

THEOREM 4.1. (i) If

$$(4.10) \quad x_n \leq n/e + K, \quad n > N, \quad K \text{ a constant},$$

then $\Omega_n = o(1)$ as $n \rightarrow \infty$, uniformly as to θ .

(ii) If $\{x_n\}$ satisfies (4.10), then (1.6) will sum a Fourier series whenever Borel's mean sums the series.

Clearly the second part of the theorem is an immediate consequence of the first part. Writing

$$D_n = \exp(-x_n + x_n e^{it}) \int_{C_1} s^n e^{-s} / n! ds,$$

we have

$$\begin{aligned}D_n &= \exp(-x_n + x_n e^{it} + i(n+1)t) \int_{x_n}^0 w^n \exp(-\omega e^{it}) / n! dw, \\ |D_n| &\leq \int_0^{x_n} \exp(x_n \cos t - x_n - w \cos t) w^n / n! dw.\end{aligned}$$

Observing that $x_n \cos t - x_n - w \cos t \leq 0$ for $0 \leq w \leq x_n$ and for all t , we have

$$|D_n| \leq x_n^{n+1} / (n+1)! \approx (ex_n)^{n+1} / (n+1)^{n+1} (2\pi)^{1/2} (n+1)^{1/2}.$$

Hence if

$$(4.11) \quad \{ex_n/(n+1)\}^{n+1} = o((n+1)^{1/2}) \text{ as } n \rightarrow \infty,$$

it follows that $D_n = o(1)$ as $n \rightarrow \infty$, uniformly as to t . Inasmuch as (4.10) implies (4.11), the theorem is proved.

By virtue of the theorem just demonstrated, we note that the many results concerning the application of Borel's method to Fourier series, which have been derived in the literature, can be transformed into theorems for the triangular method (1.6).⁹

5. We conclude this paper with two results for the transform (1.6), which are known to exist for Borel summability.

THEOREM 5.1. *The summability method of arithmetic means is not included in the method (1.6) for any sequence $\{x_n\}$.*

Proof. For certain sequences $\{x_n\}$, the theorem is immediate from the results of the preceding section on Fourier series. However, we use here a general method invoking the Silverman-Toeplitz conditions for regularity. Let

$$t_n = e^{-x_n} \sum_{\nu=0}^n s_\nu x_n^\nu / \nu!, \quad \sigma_n = \{s_0 + s_1 + \cdots + s_n\} / (n+1), \quad n \geq 0.$$

Clearly

$$\begin{aligned} t_n &= e^{-x_n} \sum_{\nu=0}^n \{(\nu+1)\sigma_\nu - \nu\sigma_{\nu-1}\} x_n^\nu / \nu!, & \sigma_{-1} &= 0 \\ (5.1) \quad t_n &= e^{-x_n} \sum_{\nu=0}^{n-1} (\nu+1)\sigma_\nu \{x_n^\nu / \nu! - x_n^{\nu+1} / (\nu+1)!\} \\ &\quad + e^{-x_n} x_n^n (n+1)\sigma_n / n! \equiv \sum_{\nu=0}^n a_{n\nu} \sigma_\nu, \end{aligned}$$

where

$$(5.2) \quad a_{n\nu} = e^{-x_n} (\nu+1) \{x_n^\nu / \nu! - x_n^{\nu+1} / (\nu+1)!\},$$

$$\nu \leq n-1, \quad a_{nn} = e^{-x_n} (n+1) x_n^n / n!.$$

In order that the transformation of $\{\sigma_\nu\}$, defined by (5.1), be regular, it is necessary that

$$\begin{aligned} (a) \quad \sum_{\nu=0}^n a_{n\nu} &= e^{-x_n} \left\{ \sum_{\nu=0}^{n-1} (\nu+1) (x_n^\nu / \nu! - x_n^{\nu+1} / (\nu+1)!) + (n+1) x_n^n / n! \right\} \\ &= e^{-x_n} \sum_{\nu=0}^n x_n^\nu / \nu! \rightarrow 1, & n &\rightarrow \infty. \end{aligned}$$

$$(b) \quad q_n = \sum_{\nu=0}^n |a_{n\nu}| = e^{-x_n} \left\{ \sum_{\nu=0}^{\mu-1} (\nu+1) |x_n^\nu/\nu! - x_n^{\nu+1}/(\nu+1)!| \right. \\ \left. + (n+1)x_n^n/n! \right\} = O(1) \text{ as } n \rightarrow \infty.$$

Theorem 2.1 shows that (a) is satisfied if $\{x_n\}$ satisfies (2.5). In fact, we shall admit those sequences $\{x_n\}$ satisfying (2.3) with $\rho > -\infty$. Then for some k ,

$$(5.3) \quad x_n < n + k\sqrt{n}, \quad n \geq 1.$$

On the contrary, (b) cannot be fulfilled if $\{x_n\}$ satisfies (5.3). To prove this, we consider two cases.

(i) Suppose $n \leq x_n < n + k\sqrt{n}$. Then

$$q_n = -e^{-x_n} \sum_{\nu=0}^{n-1} (\nu+1) \{x_n^\nu/\nu! - x_n^{\nu+1}/(\nu+1)!\} + e^{-x_n} (n+1)x_n^n/n!$$

$$q_n = 2e^{-x_n} (n+1)x_n^n/n! - e^{-x_n} \sum_{\nu=0}^n x_n^\nu/\nu! > 2e^{-x_n} n x_n^n/n! - 1.$$

It is easy to see that the function $e^{-x}x^n$, $x, n > 0$, attains its maximum when $x = n$ and that it is a monotone decreasing function of x , $x > n$. Therefore

$$q_n > 2e^{-n-kn^{1/2}} \cdot n \cdot (n + kn^{1/2})^n/n! - 1 \equiv 2v_n - 1.$$

Inasmuch as

$$v_n \approx (n/2\pi)^{1/2} e^{-kn^{1/2}} (1 + kn^{-1/2})^n \approx (n/2\pi)^{1/2} e^{-k^2/2},$$

it follows that q_n is of the order $n^{1/2}$ for the case being considered.

(ii) Suppose $x_n < n$. Letting $\mu = [x_n]$, we have

$$q_n = e^{-x_n} \left\{ - \sum_{\nu=0}^{\mu-1} (\nu+1) (x_n^\nu/\nu! - x_n^{\nu+1}/(\nu+1)!) \right. \\ \left. + \sum_{\nu=\mu}^{n-1} (\nu+1) (x_n^\nu/\nu! - x_n^{\nu+1}/(\nu+1)!) + (n+1)x_n^n/n! \right\} \\ = e^{-x_n} \left\{ - 2 \sum_{\nu=0}^{n-1} (\nu+1) (x_n^\nu/\nu! - x_n^{\nu+1}/(\nu+1)!) + \sum_{\nu=0}^n x_n^\nu/\nu! \right\} \\ = 2e^{-x_n} (\mu+1)x_n^\mu/\mu! - 2e^{-x_n} \sum_{\nu=0}^{\mu} x_n^\nu/\nu! + e^{-x_n} \sum_{\nu=0}^n x_n^\nu/\nu! \\ > 2e^{-x_n} \mu \cdot x_n^\mu/\mu! - 2.$$

Recalling that $\mu \leq x_n < \mu + 1$, and using the monotone nature of the function $e^{-x}x^\mu$ for $x > \mu$, we obtain

$$q_n > 2e^{-\mu-1} \cdot \mu \cdot (\mu + 1)^\mu / \mu! - 2 \equiv 2w_\mu - 2.$$

Inasmuch as $w_\mu \approx (\mu/2\pi)^{1/2}$, it follows that q_n is of the order $x_n^{1/2}$ for this case.

Combining the results of (i) and (ii), we see that (b) and (5.3) are incompatible. Therefore (a) and (b) cannot be satisfied with the same sequence $\{x_n\}$, thus completing the proof of the theorem.

THEOREM 5.2. *The method (1.6) sums the sequence $s_\nu = (-1)^\nu$, $\nu = 0, 1, 2, \dots$, to the value zero for all $\{x_n\}$.*

Proof. Writing $e^{x_n}t_n = \sum_{\nu=0}^n (-1)^\nu x_n^\nu / \nu! = K(-x_n)$, and applying (3.8),

we obtain

$$\begin{aligned} t_n &= e^{-x_n}/n! \int_0^\infty (t - x_n)^n e^{-t} dt = e^{-2x_n}/n! \int_{-x_n}^\infty u^n e^{-u} du, \\ t_n &= e^{-2x_n}/n! \int_0^\infty u^n e^{-u} du + (-1)^n (e^{-2x_n}/n!) \int_0^{x_n} u^n e^{-u} du, \\ t_n &\equiv e^{-2x_n} + (-1)^n r_n. \end{aligned}$$

Using the fact that the maximum of the function $u^n e^{-u}$ occurs at $u = n$, we have

$$\begin{aligned} r_n &= e^{-2x_n}/n! \int_0^{x_n} e^{2u} u^n e^{-u} du \leq e^{-2x_n} n^n e^{-n}/n! \int_0^{x_n} e^{2u} du \\ &\leq K n^{-1/2} e^{-2x_n} (e^{2x_n} - 1) = o(1), \quad n, x_n \rightarrow \infty. \end{aligned}$$

Therefore $t_n = o(1)$, $n \rightarrow \infty$, which was to be proved.

Theorem 5.2 shows that the method (1.6) is not included in convergence, for any sequence $\{x_n\}$.

OHIO STATE UNIVERSITY.

NOTES.

1. O. Szász, "Some new summability methods with applications," *Annals of Mathematics*, vol. 43 (1942), pp. 69-83.

2. In the paper referred to in note 1, the problem was considered for Abel and related summability methods.
3. See (i) M. Kac, "Note on the partial sums of the exponential series," *Univ. Nac. Tucumán, Revista A.* 3 (1942), pp. 151-153, *Mathematical Reviews*, vol. 4 (1943), p. 194; (ii) J. V. Uspensky, *Introduction to Mathematical Probability*, New York, 1937; (iii) Pólya and Szegő, *Aufgaben und Lehrsätze aus der Analysis*, Berlin, vol. 1 (1925), p. 80.
4. R. G. Cooke, "A note on lower semi-matrices," *Journal of the London Mathematical Society*, vol. 14 (1939), pp. 154-157, has also considered the problem of analytic continuation with such triangular summability methods.
5. For the definition of the Borel polygon, see (i) E. Borel, *Leçons sur les séries divergentes*, Paris, 1928; (ii) *Introduction to the Theory of Divergent Series*, lithoprinted lectures by O. Szász, written by J. Barlaз, Ann Arbor, 1944.
6. We took $x_n = o(n)$ so that $R_n \rightarrow 0$ for all z . Clearly $R_n \rightarrow 0$ for certain z even if x_n is not $o(n)$.
7. See the books referred to in note 5.
8. Referring back to note 6, we see that it is possible to have analytic continuation even if x_n is not $o(n)$. However, the region would be smaller than the polygon.
9. For literature concerning the application of Borel's method to Fourier series, see (i) G. H. Hardy, "Remarks on some points in the theory of divergent series," *Annals of Mathematics*, vol. 36 (1935), pp. 167-181; (ii) G. H. Hardy and J. E. Littlewood, "Some new convergence criteria for Fourier series," *Annali della Scuola Normale di Pisa*, vol. 3 (1934), pp. 43-62; (iii) C. N. Moore, "On the application of Borel's method to the summation of Fourier series," *Proceedings of the National Academy of Sciences*, vol. 11 (1925), pp. 284-287; (iv) L. Lorch, "The Lebesgue constants for Borel summability," *Duke Mathematical Journal*, vol. 11 (1944), pp. 459-467.

TOPOLOGICAL RINGS.*

By IRVING KAPLANSKY.

1. **Introduction.** The paper consists of two sections which may, for the most part, be read independently. Part I is devoted mainly to the study of compact rings. We first give some preliminary results, including a summary of Jacobson's [11] theory of the radical in a form convenient for our use, some remarks on "*Q*-rings" (rings in which elements near zero are quasi-regular), and an abstract theory of boundedness. The structure theorems for compact rings are given in Theorems 14-20. A noteworthy feature is the extent to which compactness serves as a substitute for the classical chain conditions.

In Part II the ring of continuous functions from a topological space X to a topological ring A is studied. When X is totally disconnected, the theory of the maximal, prime, or primitive ideals can be worked out satisfactorily with mild restrictions on A . We also give a generalization of the classical case where A is the ring of real or of complex numbers. The paper concludes with a study of two subrings: the bounded functions and those vanishing at a point.

PART I. The Structure of Compact Rings.

2. ***Q*-rings.** By a *topological ring* A we shall mean an (associative) ring which is also a Hausdorff space, such that $a \mapsto b$ and ab are continuous functions of a and b . The additional assumption of the continuity of the inverse, whenever it exists, will be considered below.

In a *metric ring* there is associated with every element a real number $|a|$ such that $|0| = 0$, $|a| > 0$ for $a \neq 0$, $|-a| = |a|$, $|a + b| \leq |a| + |b|$, $|ab| \leq |a| |b|$. If in the last statement equality holds, we say that the ring has a *valuation*. Evidently a ring with valuation can have no divisors of zero.

In A we introduce the operation \circ defined by $x \circ y = x + y + xy$. It is readily verified to be an associative operation with 0 acting as an identity

* Received September 18, 1946.

element. In general A does not form a group or even a semigroup under o , for the cancellation law may fail.

An element is said to be right quasi-regular (r. q. r.) if there exists an element y with $xoy = 0$; the latter is a right quasi-inverse of x . Left quasi-inverses and quasi-regularity are similarly defined. As is well known for associative systems, if x has a right quasi-inverse y and a left quasi-inverse z , then $y = z$; we then call y the quasi-inverse of x and say that x is quasi-regular (q. r.).

The following related result can be proved: if x has a *unique* right quasi-inverse y , then y is also a left quasi-inverse. For it follows from the identity

$$xo(yox + y) = (xoy)(2 + x)$$

that $yox + y$ is a right quasi inverse of x . Hence $yox + y = y$, $yox = 0$.

A right ideal is r. q. r. if it consists entirely of r. q. r. elements. The following lemma is useful.

LEMMA 1. *If a is in a r. q. r. ideal, and x is r. q. r., then $a + x$ is r. q. r.*

Proof. Suppose $xoy = 0$. Since $a + ay$ is in the right ideal generated by a , we have $(a + ay)oz = 0$. The following identity then shows that yoz is a right quasi-inverse of $a + x$:

$$(1) \quad (a + x)oyoz = (a + ay)oz + (xoy)(1 + z).$$

We define the radical R of A to be the (set-theoretic) join of all r. q. r. ideals. It follows readily from Lemma 1 that R is a right ideal. To prove that R is also a left ideal, we must show that $ba \in R$ where $a \in R$, $b \in A$, and for this it will suffice to prove that $ba(c + i)$ is r. q. r. for any integer i and $c \in A$. Let $x = a(c + i)$ and let y be a right quasi-inverse of xb . The identity

$$bxo(-bx - byx) = -b(xboy)x$$

shows that bx is r. q. r.

A ring A whose radical is all of A is said to be a radical ring; one whose radical is 0 is semi-simple.

For brevity we shall call a topological ring a Q_r -ring if its r. q. r. elements form an open set. For rings with unity element this is equivalent to saying that the elements with right inverses form an open set. The following lemma provides a criterion which is often more convenient.

LEMMA 2. *A ring which has a neighborhood of 0 consisting of r. q. r. elements is a Q_r -ring.*

Proof. Let x be any r. q. r. element and y a right quasi-inverse. By taking a sufficiently small we can be sure that $a + ay$ is r. q. r., say $(a + ay)oz = 0$. Then by (1), $a + x$ is r. q. r. Hence x has a neighborhood consisting of r. q. r. elements.

If the left q. r. elements are open we shall call the ring a Q_l -ring, and if the q. r. elements are open, a Q -ring. It follows from Lemma 2 that a ring is a Q -ring if and only if it is both a Q_l -ring and a Q_r -ring. However I do not know of any Q_r -ring which is not a Q -ring.

The force of the hypothesis that a ring is a Q -ring can be assessed in part from the fact (Theorem 21) that a locally compact ring without divisors of zero is a Q -ring, and from the fact that any complete metric ring is a Q -ring. The last assertion follows from Lemma 2 and the following lemma.

LEMMA 3. *In a complete metric ring an element x with $|x| < 1$ is q. r.*

Proof. The series $-x + x^2 - x^3 + \cdots + (-x)^n + \cdots$ converges say to y , and we have $xoy = yox = 0$.

Thus the condition that a ring be a Q -ring is somewhat related to its completeness. However completeness is neither necessary nor sufficient.

(a) Example of a Q -ring which is not complete: the rationals under the ordinary topology.

(b) Example of a complete (even compact) ring which is not a Q -ring: the Cartesian direct sum of an infinite number of finite fields.

Let x' denote the quasi-inverse of x , whenever this exists. A question of some importance is whether the function $x \rightarrow x'$ is continuous. We first note the following result.

LEMMA 4. *In a Q -ring the quasi-inverse is continuous wherever defined if it is continuous at 0.*

Proof. Let U be any neighborhood of x' . There exists a neighborhood V of 0 such that $z \in V$ implies $x'oz \in U$, and a neighborhood W of 0 such that $a \in W$ implies that $a + ax'$ has a quasi-inverse z in V . Then (1), with $y = x'$, shows that the quasi-inverse of $a + x$ is $x'oz$, and the latter lies in U .

LEMMA 5. *The quasi-inverse is continuous in a complete metric ring.*

Proof. By Lemmas 3 and 4, it suffices to prove continuity at 0. Since $x' = \Sigma (-x)^n$, we have that $|x| < \text{Min}(1/2, \epsilon/2)$ implies $|x'| < \epsilon$.

We shall show below (Theorem 22) that the quasi-inverse is also continuous in a locally compact ring without divisors of 0. However to encompass

more general cases we shall in general *assume the continuity of the quasi-inverse as an additional hypothesis*. It is to be noted that in case the ring has a unity element, this axiom coincides with the assumption of the continuity of the ordinary inverse, which has often been added to the axioms for a topological ring.

With the aid of the foregoing axiom we can prove the following theorem.

THEOREM 1. *If A is a Q -ring with continuous quasi-inverse, and A_n is the ring of all n by n matrices with elements in A , then A_n is a Q -ring.*

Proof. It is understood that the topology in A_n is that of the Cartesian product, i. e., if U is a neighborhood of 0 in A , then a neighborhood of 0 in A_n consists of all matrices (a_{ij}) with $a_{ij} \in U$.

For a matrix (a_{ij}) with elements suitably near 0, we then seek a matrix (b_{ij}) with

$$(2) \quad a_{ij} + b_{ij} + \sum_{k=1}^n a_{ik}b_{kj} = 0 \quad (i, j = 1, \dots, n).$$

Keeping j temporarily fixed, this gives n equations for the n unknowns b_{1j}, \dots, b_{nj} . It is convenient to establish the existence of a solution of these equations by the following lemma.

LEMMA 6. *If c_{ij}, d_i ($i, j = 1, \dots, n$) are elements of a Q -ring within a sufficiently small neighborhood of 0, then the equations*

$$(3) \quad x_i + \sum_{j=1}^n c_{ij}x_j = d_i \quad (i = 1, \dots, n)$$

have a solution.

Proof. Let c' be the quasi-inverse of c_{11} (it exists if c_{11} is near enough to 0). We have $c'o(x_1 + c_{11}x_1) = c' + x_1$. Then on applying $c'o$ on the left of the first of the equations (3), we get an equation which is explicitly solved for x_1 . We substitute for x_1 in the remaining $n-1$ equations and obtain a system of equations of the same form as (3), with coefficients which are polynomials in the c_{ij} , d_i , and c' . By induction we may assume that these $n-1$ equations can be solved, provided their coefficients are sufficiently near 0, and this can be assured by taking the c 's and d 's sufficiently near 0. By virtue of our assumption of the continuity of the quasi-inverse, c' will be near 0 along with c_{11} .

Since the equations (2) are of the form (3), it follows from Lemma 6

that (2) can be solved for b_{ij} provided (a_{ij}) is near 0, and then the matrix (a_{ij}) has the right quasi-inverse (b_{ij}) . Hence A_n is a Q_r -ring. Similarly A_n may be proved to be a Q_l -ring, and therefore it is a Q -ring.

Remark. If A is commutative, we can rearrange the above proof so as to avoid the assumption of the continuity of the quasi-inverse. The ordinary algebraic method of elimination would be available: multiply the first of equations (3) by c_{i1} , the i -th by $1 + c_{i1}$, and subtract ($i = 2, \dots, n$).

3. The radical. We shall consider in this section questions related to the nature of the radical and in particular to whether it is closed. First we prove the following lemma.¹

LEMMA 7. *If e is an idempotent in the Q_r -ring A , and B is a right ideal dense in eA , then $B = eA$.*

Proof. B contains elements $e + x$ with x arbitrarily small, say so small that $xey = 0$. Left-multiplying by e and using $ex = x$, we obtain $xexy = 0$. Then

$$(e + xe)(e + y) = e + xexy = e.$$

Since $e + xe = (e + x)e \in B$, we have $e \in B$ and $B = eA$.

Let M be a maximal right ideal in a Q_r -ring A which has a left unity element. Then the closure of M is either M or A . But in the latter case Lemma 7 implies that $M = A$. The same argument works for a maximal ideal.² We therefore obtain the following theorem.

THEOREM 2. *In a Q_r -ring with left unity element the maximal (right) ideals are closed.*

Since³ in a ring with left unity element the radical R is the intersection of the maximal right ideals, Theorem 2 implies further that R is closed. This may be proved for Q_r -rings without unity element.

THEOREM 3. *In a Q_r -ring the radical is closed.*

Proof. Let y be an element in the closure \bar{R} of R . We can write $y = a + x$ with $a \in R$ and x r. q. r. By Lemma 1, y is r. q. r. Hence \bar{R} consists of r. q. r. elements, $\bar{R} \subseteq R$, $\bar{R} = R$.

¹ Prof. Ambrose informs me that Lemma 7 fills in a slight gap in the proof of Theorem 4.3 of [1].

² "Ideal" will always mean two-sided ideal.

³ [11, Th. 18, Cor. 2]. The proof given there actually makes use only of a left unity element.

THEOREM 4. *If the r. q. r. elements are closed, then the radical is closed.*

Proof. Again \bar{R} will consist of r. q. r. elements, whence $\bar{R} = R$.

Theorems 3 and 4 provide useful classes of rings with closed radicals. We shall now give an example of a ring with a radical which is not closed.⁴

Let C be the set of all power series

$$\alpha = a_0 + a_1x + \cdots + a_nx^n + \cdots$$

with real coefficients, such that $\sum |a_i|/i!$ converges. This sum is to be the norm of α , written $\|\alpha\|$. Elements α and β are added and multiplied in the usual manner for formal power series. It is readily verified that $\alpha + \beta$ and $\alpha\beta$ are again elements of C and that we have $\|\alpha + \beta\| \leq \|\alpha\| + \|\beta\|$, $\|\alpha\beta\| \leq \|\alpha\| \|\beta\|$, and $\|a\alpha\| = |a| \|\alpha\|$ for a real. Next, C is complete. For let $\alpha_i = \sum a_{ij}x^j$ ($i = 1, 2, \cdots$) be a sequence of power series with $\|\alpha_m - \alpha_n\| \rightarrow 0$ as $m, n \rightarrow \infty$. We have $|a_{mj} - a_{nj}|/j! \leq \|\alpha_m - \alpha_n\|$, and hence for fixed j the sequence a_{1j}, a_{2j}, \cdots converges say to a_j . The norms $\|\alpha_i\|$ are bounded, say by M . Then

$$\sum_{j=0}^n |a_j|/j! = \lim_{i \rightarrow \infty} \sum_{j=0}^n |a_{ij}|/j!$$

must also be bounded by M . Hence $\sum |a_j|/j!$ converges, and the series $\alpha = \sum a_jx^j$ is in C . Given ϵ , suppose that $\|\alpha_i - \alpha_p\| < \epsilon$ for $i \geq p$. Then $\|\alpha - \alpha_p\| \leq \epsilon$. For if not, then for some n

$$\epsilon < \sum_{j=0}^n |a_j - a_{pj}|/j! = \lim_{i \rightarrow \infty} \sum_{j=0}^n |a_{ij} - a_{pj}|/j!,$$

which is impossible.

The ring C is a real normed ring in the sense of Gelfand [7]. We next show that its radical is the ideal (x) , i. e., the set of all polynomials with constant term zero. Given any such power series $\alpha = a_1x + a_2x^2 + \cdots = x\beta$, with $\|\beta\| = a$, we have $\|\alpha^n\| \leq \|x^n\| \|\beta^n\| \leq a^n/n!$. The series $\gamma = -\alpha + \alpha^2 - \alpha^3 + \cdots$ converges since it is dominated by the series for e^a , and we have $\gamma\alpha = \alpha\gamma = 0$. Thus every element in the ideal (x) is q. r., so that the radical contains (x) . Moreover it is evident that no power series with a constant term different from zero can be in the radical. Hence the radical is (x) . (We could also verify that $\|\alpha^n\|^{1/n} \rightarrow 0$ and use Gelfand's characterization of the radical.)

⁴ The ring C was constructed in conversation with H. Rubin.

Now let D be the set of all sequences $X = (\alpha_1, \alpha_2, \alpha_3, \dots)$ of elements of C with the property that $\|\alpha_n\|$ increases no faster than a polynomial in n . Formally: for some k (depending on X) and all n we have $\|\alpha_n\| < n^k$. Elements of D are to be added and multiplied component-wise, and the topology is that of point-wise convergence. It is evident that D is a topological ring.

We define a sequence Y_n of elements of D by

$$Y_n = (x, 2x, \dots, nx, 0, 0, \dots).$$

These elements approach the element $Y \in D$ which has nx in the n -th component for all n . Each Y_n is clearly in the radical of D ; but Y is not for it is not even quasi-regular. Its quasi-inverse would have to be an element $Z = (\beta_1, \beta_2, \dots, \beta_n, \dots)$ with β_n the quasi-inverse of nx . But $\beta_n = -nx + n^2x^2 - n^3x^3 \dots$ so that $\|\beta_n\| = e^n - 1$. Hence Z is not in D . Thus the radical elements Y_n approach an element not in the radical; the radical of D is not closed.

To conclude this section we shall note some generalizations of theorems which are well known in the discrete case.

LEMMA 8. *If in a Q_r -ring the set $\{x + ax\}$, for fixed a and variable x , is dense, then a is r. q. r.*

Proof. We have $x + ax = -a + y$ with y arbitrarily small. Hence we may take y r. q. r., say $yoz = 0$. Then $aoxoz = yoz = 0$.

THEOREM 5. *A Q_r -ring with no proper closed right ideals is either a radical ring or a division ring.*

Proof. If A is not a radical ring, suppose that a is not r. q. r. By Lemma 8 the right ideal $\{x + ax\}$ cannot be dense in A . Hence $x + ax = 0$ for all x and $-a$ is a left unity element. Take a maximal right ideal M in A . If $M \neq 0$, M is both dense in A and closed by Theorem 2, a contradiction. Hence A has no proper right ideals at all. Such a ring is known to be a division ring.

We remark that the hypothesis that A is a Q_r -ring cannot be dropped in Theorem 5. An example is furnished by the ring of all rational numbers which can be written in the form ap^n with a an integer prime to p and n any integer, under the p -adic topology. This ring has no proper closed ideals but it is not a field.

THEOREM 6. *If a Q -ring has no proper closed ideals, its center is either a radical ring or a field.*

Proof. Suppose that the center contains an element a which is not q. r. It follows, as in the proof of Theorem 5, that $-a$ is a unity element 1. Now for any $b \neq 0$ in the center, bA is dense, so that $bx = 1 + y$ with $yo z = 0$, $bx(1 + z) = 1$, whence the center is a field.

If an element a in the center is q. r., then its (unique) quasi-inverse b is also in the center. For $xoa = aox$ for every x , and applying b left and right we obtain $box = xob$ whence b is in the center. It follows that if the center consists of q. r. elements, it is a radical ring.

4. Boundedness. Following Shafarevich [16] we shall say that a subset S of a topological ring A is *right bounded* if for any neighborhood U of 0 there exists a neighborhood V such that $V \cdot S \subset U$.⁵ Left boundedness is similarly defined, and a set is bounded if it is both left and right bounded. We shall further say that A is locally (right) bounded if there exists in A an open set which is (right) bounded.

It is evident that a ring which has a neighborhood system consisting of (right) ideals is (right) bounded. This remark suggests the following example of a right bounded ring which is not left bounded; the ring of infinite matrices over a field having only a finite number of non-zero elements in each row, under the finite topology (a neighborhood of 0 consists of all matrices with first n rows 0). This ring is right bounded but not left bounded; in fact it is not even locally left bounded.

It is not true conversely that a bounded ring has ideal neighborhoods of 0; for a counter-example take any connected abelian group and make it a ring by defining all products to be 0. We can however prove the following partial converse.

LEMMA 9. *If a (right) bounded ring has a system of group neighborhoods of 0, then it has a system of (right) ideal neighborhoods of 0.*

Proof. Suppose that A is right bounded. Given a neighborhood U , let V be a group neighborhood with $V + V \subset U$, and W a group neighborhood with $W \subset V$, $WA \subset V$. Then $W + WA$ is an open right ideal contained in U . If A is bounded we use a similar argument for $W + WA + AW + WAW$.

⁵ We use $A \cdot B$ to mean the set of all products ab , AB to mean the set of all sums of such products.

In case A is a metric ring, boundedness with respect to the metric implies boundedness as we have defined it, but not necessarily conversely: take the reals with all products equal to 0. However in a ring with a valuation the two notions coincide.

We list some further properties of boundedness which may be readily verified by the reader.

- (1) Any discrete ring is bounded.
- (2) Any subset of a bounded set is bounded.
- (3) The closure of a bounded set is bounded.
- (4) Any convergent sequence is bounded.
- (5) The set-theoretic union of a finite number of bounded sets is bounded.
- (6) If S and T are bounded, so are $S + T$ and $S \cdot T$.
- (7) If $a_i \rightarrow 0$ and $\{b_i\}$ is bounded, then $a_i b_i \rightarrow 0$.
- (8) The Cartesian sum of bounded rings is bounded.

The concept of boundedness derives much of its importance from the following result.

LEMMA 10. *Any compact set is bounded.*

Proof. Given a neighborhood U and any point x in the compact set K , we may find neighborhoods $V(x)$, $W(x)$ of x and 0 such that $V(x) \cdot W(x) \subset U$. A finite number of the V 's cover K and if X is the intersection of the corresponding W 's we have $KX \subset U$.

We shall now prove several theorems concerning bounded rings. It of course follows from Lemma 10 that these results extend at once to compact rings.

THEOREM 7. *The radical of a right bounded Q_r -ring is open.*

Proof. Let U be a neighborhood of 0 consisting of r. q. r. elements, and V a neighborhood with $V \cdot A \subset U$. Then for $x \in V$ and any $a \in A$, xa is r. q. r. Hence V is contained in the radical of A , which is therefore open.

COROLLARY 1. *A right bounded semi-simple Q_r -ring is discrete.*

COROLLARY 2. *A compact semi-simple Q_r -ring is finite.*

THEOREM 8.^a *If C is the component of 0 in a right bounded locally compact ring A , then $CA = 0$.*

^a This generalizes a result given by Otobe [14].

Proof. For any fixed element f in the character group of A , let $I(f)$ denote the set of all $a \in A$ with $f(aA) = 0$. We show that $I(f)$ is open. Choose neighborhoods U, V with $f(U) < 1/2$, $V \cdot A \subset U$. Then for $x \in V$ we have $nx \in U$ for all integers n and hence $f(nxa) = nf(xa) < 1/2$ for all $a \in A$. It follows that $f(xa) = 0$, $x \in I(f)$, and $I(f)$ is open. Since $I(f)$ is a group, it is also closed. We then have $C \subset I(f)$, $f(CA) = 0$ for all f , whence $CA = 0$.

THEOREM 9. *A locally compact bounded ring with no proper closed ideals is discrete.*

Proof. The component of 0 in the given ring A is a closed ideal which must be either 0 or A . Hence A is either connected or totally disconnected. In the latter case it has group neighborhoods [13, 450], hence by Lemma 9 ideal neighborhoods of 0. This is possible only if A is discrete.

If A is connected, then by Theorem 8 we have $A^2 = 0$; in effect A is a connected locally compact abelian group with no proper closed subgroups. It follows [18, 110] that A is compact. But then the character group is discrete and has no proper subgroups at all; it must be finite which means that A is finite.

The theorem [9, Lemma 1] that a compact field is finite may now be generalized as follows.

COROLLARY. *A compact ring with no proper closed ideals is finite.*

Still another generalization is the following theorem.

THEOREM 10. *A right bounded division ring is discrete.*

Proof. In a division ring A we have $1 \in U \cdot A$ for any $U \neq 0$. Hence A can be right bounded only if it is discrete.

5. Compact semi-simple rings. We shall say that an element x is nilpotent if $x^n \rightarrow 0$. (In the present context there will be no confusion with purely algebraic nilpotence which of course corresponds to the discrete case.) A ring is a nil-ring if all its elements are nilpotent, and it is nilpotent if for any neighborhood U of 0 there exists N such that $R^n \subset U$ for $n > N$.

A nilpotent ideal is not necessarily in the radical: in the ring of integers under the p -adic topology the radical is 0, but the ideal (p) is nilpotent. However under suitable further assumptions the result is true.

THEOREM 11. *In a Q_r -ring, if a subsequence of $\{x^n\}$ approaches 0, then x is r. q. r.*

Proof. For some large n we have $(-x)^n$ r. q. r., say $(-x)^n oy = 0$. Set $z = [-x + x^2 - x^3 + \cdots + (-x)^{n-1}]oy$. Then $xoz = (-x)^n oy = 0$.

COROLLARY. *In a Q_r -ring the radical contains all nil right ideals.*

THEOREM 12. *If $x^n \rightarrow 0$ in a complete ring with group neighborhoods of 0, then x is q. r.*

Proof. The series $-x + x^2 - x^3 + \cdots$ converges and, as in the proof of Lemma 3, its sum is a quasi-inverse of x .

COROLLARY. *In a complete ring with group neighborhoods of 0, the radical contains all nil right or left ideals.*

For compact rings it is possible to say a good deal more. We first prove that the radical is closed.

THEOREM 13. *In a compact ring the r. q. r. elements are closed.*

Proof. Suppose that x is not r. q. r., i. e., for all y , $xoy \neq 0$. We can find neighborhoods U_x, V_y of x and y such that $0 \notin U_x o V_y$. A finite number of the V_y 's cover the ring. If U is the intersection of the corresponding U_x 's, then U is a neighborhood of x which contains no r. q. r. elements.

From Theorem 4 we obtain the desired corollary.

COROLLARY. *The radical of a compact ring is closed.*

THEOREM 14. *The radical R of a compact totally disconnected ring A is nilpotent.*

Proof. Let B be an ideal neighborhood of 0. Since a homomorphism preserves quasi-regularity, the radical of $A - B$ contains all $r + B$ with $r \in R$. Also, since $A - B$ is finite its radical is nilpotent. Hence for large n we have $R^n \subset B$, as desired.

Because of the lack of group neighborhoods of 0, for a general compact ring we can only make the following weaker assertion.

THEOREM 15. *For any neighborhood U of 0 in a compact ring there exists N such that $R \cdot R \cdots R$ (n factors) $\subset U$ for $n > N$.*

Proof. Let C be the component of 0, and V a neighborhood with $V \cdot V \subset U$. The radical $R - C$ of $A - C$ is nilpotent, whence $R^m \subset V + C$ say for $m \geq M$, and we need only take $N = 2M$.

Combining Theorems 12 and 15 we obtain the following result.

COROLLARY. *In a compact ring the radical is the union of all nil right and left ideals.*

For later use we now prove the following lemma.

LEMMA 11. *In a ring A with the descending chain condition on left ideals; the radical R contains the intersection of the maximal two-sided ideals.*

Proof. The ring $A - R$ is a semi-simple ring with descending chain condition, and so the intersection of its maximal ideals is 0. Suppose $a \notin R$; then there exists in $A - R$ a maximal ideal M not containing $a + R$. The inverse image in A of M is a maximal ideal not containing a .

We now prove the following conclusive structure theorem.

THEOREM 16. *A compact semi-simple ring is isomorphic and homeomorphic to a Cartesian direct sum of finite simple rings*

Proof. If C is the component of 0 in our ring A , we have $C^2 = 0$ by Theorem 8. Since A is semi-simple, $C = 0$ and A is totally disconnected. By Lemma 9, A has ideal neighborhoods of 0. For such an open ideal B , $A - B$ is compact and discrete and hence finite, and so contains an open maximal ideal whose inverse image in A is likewise an open maximal ideal.

Having thus proved the existence of open maximal ideals, we next show that the intersection of all of them is 0. Suppose on the contrary that $a \neq 0$ is in every open maximal ideal; the same will be true of ax for any x . Take an open ideal B . In $A - B$, the coset $ax + B$ will lie in every maximal ideal and so, by Lemma 11, will be nilpotent. We thus have $(ax)^n \in B$ for large n . By Theorem 12, a is in the radical of A , a contradiction.

Choose a fixed well-ordering $\{M_\lambda\}$ of the open maximal ideals. We select a subset as follows. Take $N_1 = M_1$, and, having chosen N_ρ for $\rho < \lambda$, take N_λ to be the first M not containing the intersection of a finite number of previously selected N 's. When the process concludes we evidently have a meet-independent [2, 64] set $\{N_\lambda\}$ with intersection 0. It is immediate that we have a one-one representation of A as a subring of the direct sum of the rings $A - N_\lambda$. Moreover the mapping is continuous: a neighborhood of 0 in the direct sum consists of all elements having 0 at a finite number of places, and the intersection of the corresponding maximal ideals is an open set mapping into it.

For any finite number of N 's we have [2, Th. 3.21] that $A - (N_{i_1} \cap \cdots \cap N_{i_k})$ is the direct sum of $A - N_{i_1}, \cdots, A - N_{i_k}$. Hence in the representation of A any finite combination of coordinates occurs, i. e.,

we have a dense subring of the direct sum. But the continuous image of a compact set is closed; hence we have the full direct sum. Since both spaces are compact, the mapping is also continuous in the reverse direction. Finally each $A - N_\lambda$ is certainly a finite ring with no proper ideals; it must be simple for if it were a zero ring A would not be semi-simple.

Remarks. 1. In the light of the revealed structure of A it is apparent that the N 's are in fact the only open (or for that matter closed) maximal ideals in A . Thus the process of selecting the N 's from the M 's can be seen in retrospect to have been vacuous.

The following are fairly direct corollaries of Theorem 16.

2. A compact semi-simple ring has a unity element.
3. A compact semi-simple ring which has the ascending chain condition on closed ideals, or the descending condition on open ideals, is finite.
4. In a compact ring left and right quasi-inverses coincide (for this is true in any ring if it is true modulo the radical).
5. In a compact ring with unity element the radical is the intersection of the open maximal ideals.
6. A compact semi-simple ring satisfying the second axiom of countability is the direct sum of a countable number of finite simple rings.
7. A compact primitive ring (in the sense of Jacobson [11]) is finite.

6. Compact rings with radical. Since a finite simple ring is a matrix ring over a finite field, Theorem 16 assures us that a compact semi-simple ring has an ample supply of idempotents. We now require a device for transferring these idempotents to a ring with radical.

LEMMA 12. *Let A be a compact ring, B a nilpotent ideal in A , and H the homomorphism from A to $A - B$. Suppose that in $A - B$ we have a well-ordered set of idempotents $\{f_\rho\}$ such that $f_\alpha f_\beta = f_\beta f_\alpha = f_\alpha$ for $\alpha \leq \beta$, and for λ a limit ordinal, $f_\lambda = \lim_{\rho < \lambda} f_\rho$. Then there exists in A a set of idempotents $\{F_\rho\}$ with $HF_\rho = f_\rho$, and $F_\alpha F_\beta = F_\beta F_\alpha = F_\alpha$ for $\alpha \leq \beta$.*

Proof. Suppose that F_ρ has been found with the desired properties for $\rho < \lambda$.

Case I. λ not a limit ordinal. Let c be any element with $Hc = f_\lambda$. Define

$$c_1 = c - cF_{\lambda-1} - F_{\lambda-1}c + F_{\lambda-1}cF_{\lambda-1}.$$

Then

$$(4) \quad Hc_1 = f_\lambda - f_{\lambda-1}, \quad c_1F_\alpha = F_\alpha c_1 = 0 \quad (\alpha \leq \lambda - 1).$$

Define inductively $c_{n+1} = 3c_n^2 - 2c_n^3$. Then

$$(5) \quad c_{n+1}^2 - c_{n+1} = 4(c_n^2 - c_n)^3 - 3(c_n^2 - c_n)^2.$$

We have $H(c_1^2 - c_1) = 0$, i. e., $c_1^2 - c_1 \in B$; by induction it follows from (5) that $c_n^2 - c_n \in B^{2^n}$. Let d be a limit point of $\{c_n\}$; then d is an idempotent and moreover, since each c_n shares properties (4) with c_1 , so does d . On taking $F_\lambda = F_{\lambda-1} + d$ we successfully continue the induction.

Case II. λ a limit ordinal. Let F_λ be a limit point of F_ρ ($\rho < \lambda$). By continuity $HF_\lambda = f_\lambda$ and $F_\alpha F_\lambda = F_\lambda F_\alpha = F_\alpha$ for $\alpha \leq \lambda$, as desired.

We shall define a ring to be (*completely*) *primary* if it has a unity element, and modulo its radical is simple (a division ring). It is to be observed that for commutative rings the two notions coincide. We can prove the following structure theorem.⁷

THEOREM 17. *A commutative compact ring is the Cartesian direct sum of a radical ring and primary rings.*

Proof. Let A be the compact ring, R its radical, C the component of 0. We have that $C^2 = 0$ (Theorem 8) and the radical $R - C$ of $A - C$ is nilpotent (Theorem 14). By two successive applications of Lemma 12 we find in A an idempotent e mapping on the unity element of $A - R$. The Pierce decomposition relative to e expresses A as the direct sum of a radical ring and a ring with unity element. Our further discussion may therefore be confined to the latter, i. e., we assume that A has a unity element and is therefore totally disconnected.

The ring $A - R$ is a direct sum of fields, say with generating idempotents $\{e_\rho\}$. Define $f_\lambda = \sum_{\rho < \lambda} e_\rho$, apply Lemma 12, and finally define $E_\lambda = F_{\lambda+1} - F_\lambda$. The E 's form a set of mutually orthogonal idempotents,

⁷ For rings with ascending chain condition, an analogous theorem has been given by v. Dantzig [4, 213]. Jacobson's theorem [10, 442] that a compact ring is a direct sum of p -rings is both more and less general, since the ring need not be commutative while on the other hand p -rings are in general not primary.

and each AE_λ is a compact primary ring. The correspondence $a \rightarrow \{aE_\lambda\}$ maps A into the direct sum of the AE_λ 's. It is a one-one correspondence, since an element which annihilates each E_λ also annihilates each F_λ , and hence also the final F which is 1. The mapping is continuous, for if the directed set $\{a_\alpha\}$ converges to a , then $a_\alpha e$ converges to ae for any element e . Any finite combination of coordinates arises; hence we get the full direct sum. Finally since both spaces are compact, the mapping is continuous in the reverse direction.

It is well known that Theorem 17 is not valid for non-commutative rings, even in the discrete case. One must be content with a decomposition into left ideals, or into an additive direct sum of primary rings [5, 16-18]. Analogues of these two decompositions can be proved without difficulty for compact rings. It is also possible to prove the following theorem.⁸

THEOREM 18. *A compact primary ring is a matrix ring over a compact completely primary ring.*

We require the following lemma.⁹

LEMMA 13. *Let e be an idempotent in a compact ring A with radical R . Then eRe is the radical of eAe .*

Proof. In any ring we always have that eRe is contained in the radical of eAe . For given $r \in eRe$, we need only show that for any a , rae is r. q. r. in eAe . But if $(rae)ob = 0$, then $(rae)o(ebe) = 0$.

Conversely eRe is a nil ideal which by Theorem 15 is contained in the radical of eAe .

Proof of Theorem 18. Let A be the ring, R its radical; $A - R$ is a matrix ring over a finite field, say with matrix units e_{ij} ($i, j = 1, \dots, n$). By Lemma 12, we have in A orthogonal idempotents E_1, \dots, E_n mapping on e_{11}, \dots, e_{nn} .

We next show that there exist elements $a, b \in A$ such that $aE_1b = E_i$. The proof of this merely requires a re-examination of the construction used in Lemma 12. Let x, y be any elements mapping on e_{i1}, e_{1i} respectively. If necessary we can replace x by E_ix , y by yE_i , and so we may assume $E_ix = x$,

⁸ Our proof is actually valid for any complete topological ring whose radical is nilpotent, and which modulo the radical is a simple ring with minimum condition.

⁹ Lemma 13 will hold under any hypothesis which assures us that the radicals of A and eAe are the union of all nil-ideals, for example A may be a normed ring.

$yE_i = y$. We take xE_1y to be the c_1 of Lemma 12. Then $c_2 = 3c_1^2 - 2c_1^3 = xE_1y_2$ where $y_2 = (3 - 2yxE_1)yxE_1y$, and in general $c_n = xE_1y_n$. Now c_n approaches an idempotent E , and if b is a limit point of $\{y_n\}$, we have $E = xE_1b$. Evidently $bE_i = b$ and so $EE_i = E_iE = E$. Also E and E_i both map on e_{i1} . Hence $E - E_i$ is an idempotent contained in R , whence $E = E_i$.

Now with $E_i = aE_1b$, define $E_{1i} = E_1b$, $E_{i1} = aE_1$, $E_{ij} = E_{i1}E_{1j}$. Then $E_{i1}E_{1i} = E_{ii} = E_i$. Also $E_{1i}E_{i1}$ is an idempotent mapping on e_{11} and is invariant under multiplication by E_1 . Hence $E_{1i}E_{i1} = E_1$. It then follows that $E_{ij}E_{ki} = \delta_{jk}E_{ii}$.

Consider the right ideal E_1A . The mapping $a \rightarrow ax$ gives for every $x \in A$ an endomorphism of E_1A . The correspondence is one-one, for if $E_1Ax = 0$, then $E_ix = 0$ and $x = 0$, since $1 = E_1 + \dots + E_n$. Let T be an endomorphism of E_1A commuting with all these right multiplications, i. e., $(E_1a)Tx = (E_1ax)T$ for all a and x . Then $E_1T = (E_1T)E_1 = E_1bE_1$, say. Also $(E_1a)T' = (E_1T)a = (E_1bE_1)(E_1a)$; in short, T is a left multiplication by an element in E_1AE_1 .

We now regard E_1A as a left vector space over E_1AE_1 . It is easy to see that $E_1, E_{12}, \dots, E_{1n}$ form a basis for this vector space; for

$$E_1a = E_1a \cdot \sum E_i = \sum (E_1aE_{i1}E_1)E_{1i},$$

while $\sum a_iE_{1i} = 0$ implies $a_iE_1 = 0$ on right multiplication by E_{i1} .

We have represented A as a set of endomorphisms on this vector space. We get all endomorphisms, since E_{ij} sends E_{1i} into E_{1j} and the other basis elements into zero. By Lemma 13, E_1AE_1 is completely primary and since it is closed it is also compact.

It remains finally to identify the topology of A with that of the matrix ring. A neighborhood U of 0 in the latter consists of matrices with elements in a neighborhood V of 0 in E_1AE_1 . Take a neighborhood W in A such that $E_{1i}xE_{j1} \in V$ for all i, j and all $x \in W$. Then the matrix representing x will lie in U . This proves continuity of the mapping one way; the reverse direction follows since both spaces are compact.

We can also establish the following result.

THEOREM 19. *A compact ring with no divisors of zero is either a radical ring or completely primary.*

Proof. In a ring with no divisors of zero there is at most one idempotent—the unity element. It follows from Lemma 12 that the ring modulo its radical is either zero or a division ring.

A commutative primary ring has a radical consisting precisely of all non-units, and will thus be a local ring if the ascending chain condition is fulfilled. We shall now give an alternative sufficient condition for this purpose.

THEOREM 20. *Let A be a compact commutative primary ring with radical R . If R^2 is open, then A is a local ring and has the latter's natural topology.*

Proof. The ring $R - R^2$ is compact and discrete and hence finite. Let a_1, \dots, a_n be any representatives in R of these cosets, and let $B = (a_1, \dots, a_n)$. Clearly $B \subseteq R$; we assert that $R = B$. Any element in R is expressible as a sum of elements in B and R^2 , and more generally one in R^i as a sum of elements in $R^{i-1}B$ and R^{i+1} . Since for any neighborhood U , $R^n \subset U$ for large n , it follows that $r \in R$ can be written $r = \sum_{i=1}^{\infty} r_i$ with r_i in $R^{i-1}B$. Write $r_i = \sum c_{ij}a_j$ ($c_{ij} \in R^{i-1}$) and we see that $\sum c_{ij}$ converges say to c'_j , $r = \sum c'_ja_j \in B$.

We have thus shown that R has a finite basis. It now follows from a result due to Cohen [3, Th. 3] that A is a local ring. Since the residue class field is finite, A is compact in the local ring topology (this topology is defined by taking $\{R^n\}$ to be a system of neighborhoods of 0). The mapping

$$A(\text{local ring topology}) \rightarrow A(\text{original topology})$$

is continuous since any neighborhood of 0 contains some R^n . Since A is compact in both topologies, the mapping is a homeomorphism.

We shall now give an example of a compact integral domain which is not a local ring. Let A_m ($m = 1, 2, \dots$) denote the ring of all formal power series in x_1, \dots, x_m with coefficients in some finite field. For $m < n$ there is a natural homeomorphism θ_{mn} of A_n into A_m obtained by simply suppressing the variables x_{m+1}, \dots, x_n . Relative to the projections θ_{mn} , $\{A_m\}$ forms an inverse system [6, § 20]. Each A_m is a compact integral domain and so the limit ring is again a compact integral domain. However A is not a local ring since, for example, $(x_1) \subset (x_1, x_2) \subset (x_1, x_2, x_3) \subset \dots$ is an ascending chain of ideals. It may also be observed that R^2 is not open since any neighborhood of 0 contains $x_n \notin R^2$ for large n .

We shall conclude this section with some remarks on locally compact rings.

LEMMA 14. *A locally compact ring without divisors of zero is either connected or totally disconnected.*

Proof. Let C be the component of 0. C is a locally compact connected ring without divisors of zero, hence [9] C is either 0 or a division algebra over the reals. In the latter event let 1 be the unity element of C . Then for any $a \in A$, $1 \cdot a = c \in C$. But then $1(a - c) = 0$, $a = c$.

LEMMA 15. *Let x be an element in a ring with the second axiom of countability and without divisors of zero such that the closure of $\{x^n\}$ is compact. Then either $x^n \rightarrow 0$ or $\{x^n\}$ has a subsequence approaching 1.*

Proof. Suppose $x^{n(i)} \rightarrow a \neq 0$. A subsequence of $\{x^{n(i+1)-n(i)}\}$ approaches y and $ay = a$. Then for any z , $a(yz - z) = 0$. If $a \neq 0$, $yz = z$ and similarly $zy = z$, $y = 1$.

LEMMA 16. *In a locally compact ring with the second axiom of countability and without divisors of zero there exists a neighborhood of 0 consisting of nilpotent elements.*

Proof. Let U be a compact neighborhood of 0; if A has a unity element 1 we may assume $1 \notin U$. Let V be a neighborhood of 0 with $V \subset U$, $VU \subset U$. Then for $x \in V$, $x^n \in U$. By Lemma 15, $x^n \rightarrow 0$.

LEMMA 17. *For a locally compact totally disconnected ring A with the second axiom of countability the following three statements are equivalent:*

- (1) A is a Q -ring,
- (2) A is a Q_r -ring,
- (3) A has a neighborhood of 0 consisting of nilpotent elements.

Proof. That (1) implies (2) is obvious, and that (3) implies (1) follows from Theorem 14. To prove that (2) implies (3) we proceed as in Lemma 15. When we reach $ay = a$ we may assume that $-y$ is r. q. r., say $-yoz = 0$. Left-multiplying by a , we obtain $ay = 0$, $a = 0$.

THEOREM 21. *A locally compact ring with the second axiom of countability and without divisors of zero is a Q -ring.*

Proof. If A is connected it is a division ring, and any division ring is a Q -ring since -1 is the only element which is not q. r. If A is totally disconnected, the theorem follows from Lemmas 16 and 17.

For the case of a division ring the following result was proved by Otake [15].¹⁰

THEOREM 22. *In a locally compact ring with the second axiom of countability and without divisors of zero the quasi-inverse is continuous.*

Proof. This is true for the reals, complexes, or quaternions, so we assume that A is totally disconnected. By Lemma 4 and Theorem 21 it suffices to prove continuity at 0. By taking x in the neighborhood V of the proof of Lemma 16, we have that its quasi-inverse $-x + x^2 - x^3 \cdots$ is in U .

PART II. Rings of Functions.

7. Functions on totally disconnected spaces. Let $C(X, A)$ denote the ring of continuous functions from a topological space X to a topological ring A . We shall consider three ways of topologizing $C(X, A)$, listed in order of increasing strength:

p : pointwise convergence,

k : uniform convergence on compact subsets of X ,

u : uniform convergence on all of X .

Of course p and k coincide if S is discrete, and k and u coincide if X is compact. When desirable we shall indicate the choice of topology by the notation $C(X, A, p)$, etc.

Some elementary facts about $C(X, A)$ are assembled in the following theorem.

THEOREM 23. *Suppose that the quasi-inverse in A is continuous where defined. $C(X, A)$ can be a Q -ring only if A is a Q -ring, and if the latter is, so is $C(X, A, u)$ but not in general $C(X, A, p$ or $k)$. The radical of $C(X, A)$ is $C(X, R)$ where R is the radical of A ; it is closed under the p -topology (or any stronger topology) if and only if R is closed. $C(X, A, u)$ is complete if and only if A is. If X is locally compact, $C(X, A, k)$ is also complete.*

¹⁰ The elements with quasi-inverses form a group under \circ which is complete in a suitable metric. Hence Theorem 22 can also be derived from a result due to Montgomery: "Continuity in topological groups," *Bulletin of the American Mathematical Society*, vol. 42 (1936), pp. 879-882. (I am indebted to the referee for this remark.)

Proof. Let $f \in C(X, A)$ be such that $f(x)$ is q.r. for all x . There is a unique function g with $f(x)og(x) = 0$ for all x . By the continuity of the quasi-inverse, g is continuous. It follows that if A is a Q -ring, so also is $C(X, A, u)$. Under the p - or k -topology, neighborhoods of 0 will in general contain functions which assume values not q.r.

We leave to the reader the remainder of the proof of Theorem 23.

The connection between the structure of $C(X, A)$ and that of X and A has been widely studied for the case where A is the reals or complexes [8], and for the case where X is totally disconnected and A is the field of integers mod 2 [17], or more generally a discrete division ring [12]. We shall proceed to study the case where X is totally disconnected under less restrictive assumptions on A ; however we shall assume that A has a unity element.

We shall use the following terminology suggested by E. Hewitt: an ideal I in $C(X, A)$ is *free* if for every $x \in X$ there exists $f \in I$ with $f(x) = 1$; otherwise I is *fixed*.

THEOREM 24. *If X is totally disconnected and compact and A is a Q -ring with unit and inverse continuous where defined, then every proper right ideal in $C(X, A)$ is fixed. A maximal (right or two-sided) ideal M has the following form: for some $x \in X$ and maximal right or two-sided ideal N in A , M consists of all f with $f(x) \in N$.*

Proof. Suppose that I is a free right ideal, and let $f \in I$ be a function with $f(x) = 1$. In a suitable neighborhood U_x of x , which we may take to be open and closed, we have that f^{-1} exists. Define a function g as follows: $g = f^{-1}$ in U_x and is 0 otherwise. The function g is continuous and so $h = fg \in I$; h is the characteristic function of U_x . A finite number of the U 's cover X ; by taking intersections and reselecting the corresponding h 's, we may suppose that we have disjoint open and closed sets U_1, \dots, U_n covering X , with characteristic functions h_1, \dots, h_n in I . Then $h_1 + \dots + h_n = 1$ is also in I and $I = C(X, A)$. The final statement of the theorem follows from the fact that quite generally any fixed maximal ideal has the designated form.

THEOREM 25. *Suppose that $C(X, A)$ is isomorphic to $C(X_1, A_1)$ where X, X_1 are totally disconnected and compact and A, A_1 are Q -rings with unit elements. Then for any maximal ideal N in A there exists a maximal ideal N_1 in A_1 such that $A - N$ and $A_1 - N_1$ are isomorphic. If A, A_1 are simple, then they are isomorphic and X, X_1 are homeomorphic.*

Proof. Take a point $x \in X$ and construct the maximal ideal M of functions f with $f(x) \in N$. Let M_1 be the corresponding maximal ideal in $C(X_1, A_1)$. By Theorem 24 there exists in A_1 a maximal ideal N_1 such that $A - N \cong C(X, A) - M \cong C(X_1, A_1) - M_1 \cong A_1 - N_1$.

If A, A_1 are simple, then $N = N_1 = 0$, so that $A \cong A_1$. Moreover there is a one-one correspondence between points of X and maximal ideals in A . If we introduce the Stone topology into the latter, we verify by well known methods that the correspondence is a homeomorphism (see for example [8]).

Theorem 25 can be interpreted as saying that when X is compact and totally disconnected and A is simple, the algebraic properties of $C(X, A)$ characterize X completely. This is no longer true if the hypothesis of compactness is dropped. However we shall now show that the algebraic and topological properties of $C(X, A, k)$ together do characterize X . At the same time we are able to weaken slightly the hypothesis on A .

THEOREM 26. *Let X be totally disconnected. Any closed (right) ideal I in $C(X, A, k)$ has the following form: for every $x \in X$ a closed (right) ideal $J(x)$ in A is prescribed and I consists of all functions f with $f(x) \in J(x)$ for all x .*

Proof. Define $K(x)$ to be the set of values in A assumed at x by functions in I and let $J(x)$ be the closure of $K(x)$. Let f be a function with $f(x) \in J(x)$ for every x ; we shall show that $f \in I$. Take a neighborhood of 0 in $C(X, A, k)$, consisting say of all functions mapping a compact set K into a neighborhood U of 0 in A . For any $x \in K$ there is a function $g \in I$ with $g(x) - f(x) \in U$, and this will extend to an open and closed neighborhood U_x of x . We may assume that g vanishes outside U_x . A finite number of the U 's cover K ; we may assume them non-overlapping. The sum $h \in I$ of the corresponding g 's then has the property that $h(x) - f(x) \in U$ for all x in K . Hence I contains functions arbitrarily close to f , and since it is closed it must contain f .

COROLLARY 1. *If in addition A has no proper closed ideals, then any closed ideal in $C(X, A, k)$ consists of all functions vanishing on some (necessarily closed) set of X . In particular a closed maximal ideal consists of all functions vanishing at a point.*

COROLLARY 2. *Let X, X_1 be totally disconnected spaces and A, A_1 rings with no proper closed ideals such that $C(X, A, k)$ is isomorphic and homeo-*

morphic to $C(X_1, A_1, k)$. Then A, A_1 are isomorphic and homeomorphic, and X, X_1 are homeomorphic.

Suppose that the maximal right ideals in A are closed (this is true by Theorem 2 if A is a Q_r -ring); their intersection is the radical R of A . It follows readily from Theorems 23 and 26 that the intersection of the closed maximal right ideals of $C(X, A, k)$ already gives the radical of that ring. This is not true for all rings, since in 3 we gave an example of a ring with non-closed radical.

It is natural to ask: what about the free maximal ideals? For locally compact spaces we can give the following result.

THEOREM 27. *Let X be a locally compact totally disconnected space and A a Q -ring with unit; and let F consist of all $f \in C(X, A)$ such that the closure of the set of points x for which $f(x) \notin R$ (the radical of A) is compact. Then a maximal right ideal in $C(X, A)$ is free if and only if it contains F .*

Proof. The ideal F is itself free; indeed it is easy to see that it is dense in $C(X, A, k)$. Thus any ideal containing F is free. Conversely suppose that M is a free maximal right ideal and that $f \in F$; we show that $f \in M$. The set of points x where $f(x) \notin R$ can be embedded (since X is locally compact) in an open compact set K . The function g , defined by $g(x) = 0$ in K and $g(x) = f(x)$ outside K , lies in the radical of $C(X, A)$; hence $g \in M$. Thus to prove $f \in M$ it will suffice to prove that $f - g \in M$, and for this it will in turn suffice to prove $k \in M$, where k is the characteristic function of K ; this last fact follows from Theorem 24.

Theorem 27 shows that the intersection of the free maximal right ideals contains F . Whether or not it is equal to F is an open question. A partial affirmative answer is afforded by the following result.

THEOREM 28. *With the same hypothesis as in Theorem 27, suppose further that either X or A is discrete. Then F is the intersection of the free maximal right ideals of $C(X, A)$.*

Proof. Suppose that on the contrary there exists a function f not in F but in every free maximal right ideal. For any $a \in A$ we shall denote by $f^{-1}(a)$ the set of $x \in X$ with $f(x) = a$. By virtue of our assumption that either X or A is discrete, $f^{-1}(a)$ is open and closed. We define two further functions g and h , distinguishing two cases.

I. $a \in R$, the radical of A . Set $g(x) = 0$, $h(x) = 1$ for $x \in f^{-1}(a)$.

II. $a \notin R$. Then there exists a maximal right ideal N in A which with a generates A , i. e., there exist elements b, c with $ab + c = 1$, and such that c does not have a right inverse. We set $g(x) = b, h(x) = c$ for $x \in f^{-1}(a)$.

The method of definition assures us that g and h are continuous functions and that $fg + h = 1$.

We show next that h is contained in some free maximal right ideal. For this, by Theorem 27, it will suffice to prove that h and F generate a proper ideal. Suppose the contrary: then there exist functions p, q , with $q \in F$, such that $hp + q = 1$. Let P denote the set of points where $1 - q$ fails to have an inverse. By the definition of F , the closure of P is compact. Hence the set P' of points where h fails to have a right inverse has compact closure. By the definition of h , P' includes all x such that $f(x) \notin R$. Hence the points x where $f(x) \notin R$ have compact closure, and $f \in F$, a contradiction.

It must then be the case that f and h together generate a proper right ideal, an absurdity in view of $fg + h = 1$.

8. Prime ideals. Let P be a prime ideal in $C(X, A)$. It is assumed that X is totally disconnected. For every $x \in X$, the values assumed by functions of P at x form an ideal $I(x)$. We assert that, with the exception of at most one point, we must have $I(x) = A$. For suppose that at the point x and y the values a and b respectively are excluded. We surround x by an open and closed set U excluding y , and form functions f, g equal to a, b on U and the complement of U respectively, and zero otherwise. Then $fg = 0$ so that either f or g is in P , a contradiction. In case A is simple, this result may be restated as follows: a prime ideal can have at most one zero. If X is compact, P has precisely one zero, as follows from Theorem 24.

In our further investigation we make the following three assumptions on A :

- (1) A has no divisors of 0,
- (2) division where possible is continuous,
- (3) for any $a \in A$ there exists a neighborhood of 0 consisting of right multiples of a .

(Examples of rings satisfying these postulates: any division ring, any discrete ring with no divisors of zero, the ring of p -adic integers, etc.)

It can be proved from these axioms that the closure Q' of a prime ideal

Q in A is prime. For suppose $ab \in Q'$ with neither a nor b in Q' . A sufficiently small neighborhood of ab consists of right multiples of a ; thus there exist elements $ac \in Q$ arbitrarily near to ab . Since $a \notin Q$, we have $c \in Q$. By continuity of division, any neighborhood of b contains an element of Q . Hence $b \in Q'$, a contradiction.

Let P again be a prime ideal in $C(X, A)$, and suppose that at the point x it takes on the values of a proper ideal I . We assert that I is a prime ideal. Suppose $f(x) = ab \in I$. In a small open and closed neighborhood U , f will continue to be a right multiple of a , and the quotient by our assumption is continuous. Let $g = a$ on U and 1 on the complement; $h = a^{-1}f$ on U and f on the complement. Then $gh \in P$, so that either g or h is in P , whence either a or b is in I .

By Theorem 26 the closure of P in $C(X, A, k)$ consists of all functions with $f(x)$ in the closure of I . We may summarize these results in the following theorem.

THEOREM 29. *Suppose that X is totally disconnected and that A satisfies the three assumptions above. Then the closure of a prime ideal in A or $C(X, A, k)$ is prime. Further the following statement is true in $C(X, A, k)$ if and only if it is true in A : a closed prime ideal is either maximal or the whole ring.*

With more drastic assumptions it is possible to prove that prime ideals in $C(X, A)$ are maximal.

THEOREM 30. *Suppose that A is a division ring and that either X or A is discrete. Then the prime ideals in $C(X, A)$ are maximal.*

Proof. Let P be a prime ideal and suppose that (P, f) is proper, $f \notin P$. Let $Z \subset X$ denote the set of zeros of the function f ; under either hypothesis Z is open and closed. Let g and h be the characteristic functions for Z and the complement of Z ; since $gh = 0$ we have either g or h in P . In the first case $g + f \in P$ is a function with no zeros, which is impossible. In the second case $f = fh$ is already in P .

We shall now give an example to show that the hypotheses of Theorem 30 cannot be light-heartedly removed.¹¹ Suppose that A is a field with a valuation and that all non-negative real numbers are assumed as values (actually a dense

¹¹ This example is based on a similar one constructed by E. Hewitt for the case where A is the reals and X the unit interval.

set would suffice). Let us take X to be compact and totally disconnected, and let $X = U_1, U_2, \dots$ be a sequence of compact open sets shrinking to the point x . Take any function f satisfying $|f| = 1/n$ on $U_n - U_{n+1}$, and another function g satisfying $|g| = e^{-n}$ on $U_n - U_{n+1}$. Then the ideal generated by g does not contain any power of f . For suppose $f^n = gh$, with $|h| \leq M$. Then on $U_n - U_{n+1}$ we have $n^{-p} \leq Me^{-n}$ which cannot be true for all n .

By Zorn's lemma, expand the ideal (g) to one which is maximal with respect to the exclusion of the powers of f . The result is a prime ideal, but not a maximal one since it lacks some of the functions vanishing at x .

9. Primitive ideals. In a ring with unit a primitive ideal [11] is the largest two-sided ideal contained in a maximal right ideal. From Theorem 24 the following result follows at once.

THEOREM 31. *Let X be compact and totally disconnected, and A a Q -ring. A primitive ideal in $C(X, A)$ consists of all functions f with $f(x) \in P$, where x is a fixed point of X and P is a primitive ideal in A .*

In particular we see that if all primitive ideals in A are maximal, the same will be true in $C(X, A)$.

Jacobson [12] has raised the following question. Let X be compact and totally disconnected, A a discrete division ring, and let C^* be a subring of $C(X, A)$ containing the constant functions and for any $x, y \in X$ a function with $f(x) \neq f(y)$. Are the primitive ideals in C^* maximal? We shall show that the answer is affirmative by proving that $C^* = C(X, A)$. Actually we can prove the following more general result (it is to be noted that in the above case $C(X, A, k)$ is discrete).¹²

THEOREM 32. *Let X be totally disconnected and A a ring having a system of ideal neighborhoods of 0. Let C^* be a subring of $C(X, A)$ containing the constant functions, and for any $x, y \in X$ a function f with $f(x) = 1$, $f(y) = 0$. Then C^* is dense in $C(X, A, k)$.*

Proof. Since we can work within a neighborhood of 0 in $C(X, A, k)$, it suffices to treat the case where X is compact. Let U be an ideal neighborhood of 0 in A , and let K be an open compact set in X , L its complement. We shall show that C^* contains a function f with $f(K) \subset 1 + U$, $f(L) \subset U$.

¹² For the case where A is a finite field, Theorem 32 was communicated to me by R. Arens.

This will complete the proof, for it is readily seen that any function in $C(X, A, k)$ can be approximated within U by a finite linear combination of such functions.

Take a fixed point $y \in L$. For any $x \in K$ there exists a function g with $g(x) = 0$, $g(y) = 1$. In a neighborhood V_x of x we have $g(V_x) \subset U$. A finite number of the V 's cover K , and the product of the corresponding g 's gives us a function h with $h(K) \subset U$, $h(y) = 1$. The function $1 - h$ takes values within U in a neighborhood W_y of y . A finite number of W 's cover L , and the product of the corresponding $(1 - h)$'s gives us the desired function f .

The space of primitive ideals of a ring, topologized after Stone, has been called the *structure space* by Jacobson. In connection with this concept, we can prove the following result.

THEOREM 33. *If X is totally disconnected and compact, the structure space of $C(X, A)$ is the Cartesian product of X and the structure space of A .*

Proof. Theorem 31, together with its obvious converse, shows that there is a one-one correspondence between the two spaces in question; moreover each of them is compact. In the direct product take the set D of all pairs (b, p) , $b \in B$, $p \in P$, where B is closed in X and P is a closed set of primitive ideals in A , say with intersection I . It will suffice to show that E , the corresponding set of primitive ideals in $C(X, A)$, is closed. Let J be the intersection of the ideals in E ; clearly J consists of all f with $f(x) \in I$ for $x \in B$. Moreover for any $y \notin B$, J contains a function with $f(y) = 1$: we may surround B by an open and closed set U excluding y , and take $f = 0$ on U , $f = 1$ on the complement of U . It follows from Theorem 31 that any primitive ideal containing J must consist of all f with $f(z) \in Q$ for a suitable point z in B and a primitive ideal Q in A . Finally we observe that Q must contain I ; since P is closed we have $Q \in P$, and it follows that E is closed.

10. Functions with values in certain division rings. Let X be a compact T_0 -space and A a division ring. It appears to be an open question whether proper ideals in $C(X, A)$ are necessarily fixed. We have seen (Theorem 24) that the answer is affirmative when X is totally disconnected. Another case where the answer is known to be affirmative is that where A is the field of reals or complexes. This latter case can be generalized as follows.

Suppose that A is a division ring admitting a continuous function $x \rightarrow x^*$ such that $xx^* + yy^* = 0$ implies $x = y = 0$. From the case $x = y$ it is evident that such a function cannot exist in a ring of characteristic 2.

However one exists in every ring of characteristic different from 2 that I have examined.

Examples. (1) Discrete. Take $0^* = 0$, $x^* = x^{-1}$ for $x \neq 0$.

(2) Formally real. Take $x^* = x$.

(3) Complete in an archimedean valuation. Then A is the reals, complexes or quaternions, and we take x^* to be the conjugate of x .

(4) Non-archimedean valuation. A possible choice is the following. For each value m take a fixed element z_m of that value. Define $0^* = 0$, $x^* = z_m^2 x^{-1}$, if $x \neq 0$ has value m .

(5) The rational numbers under the 6-adic topology. Take $0^* = 0$, and if $r = 2^m 3^n s t^{-1}$ (s, t prime to 6), $r^* = 2^{2m} 3^{2n} t^{-1}$.

In the following we shall write f^* for the function defined by $f^*(x) = [f(x)]^*$, and $Z(f)$ for the set of zeros of f . It is assumed throughout that X is a T_0 -space and that A possesses a $*$ -function.

LEMMA 18. For any functions f_1, \dots, f_n in $C(X, A)$ there exist functions g_1, \dots, g_n such that $Z(f_1 g_1 + \dots + f_n g_n) = Z(f_1) \cap \dots \cap Z(f_n)$.

Proof. By induction we may assume that we have h_1, \dots, h_{n-1} with $Z(h) = Z(f_1 h_1 + \dots + f_{n-1} h_{n-1}) = Z(f_1) \cap \dots \cap Z(f_{n-1})$. We then take $kh^* + f_n f_n^*$.

LEMMA 19. If I is a free proper right ideal in $C(X, A)$, then for every $f \in I$, $Z(f)$ is non-compact.

Proof. For any functions f_1, \dots, f_n in I , $Z(f_1) \cap \dots \cap Z(f_n)$ must be non-void by Lemma 18. Suppose that for some $f \in I$, $Z(f)$ is compact. Then the intersection of all $Z(g) \cap Z(f)$, as g ranges over I , is non-void, and I is fixed.

COROLLARY. If X is compact, every proper ideal in $C(X, A)$ is fixed.

LEMMA 20. If I is a free right ideal in $C(X, A)$ and $K \subset X$ is compact, then there exists in I a function not vanishing in K .

Proof. For any x in K we have a function $f \in I$ and a neighborhood U_x such that $Z(f) \cap U_x$ is void. A finite number of the U 's cover K , and by Lemma 18 we may build from the f 's the desired function.

COROLLARY. A closed proper right ideal in $C(X, A, k)$ is fixed.

LEMMA 21. If M is a maximal right ideal in $C(X, A)$, $f \in M$, and $Z(g) \supseteq Z(f)$, then $g \in M$.

Proof. If not we have $gh + m = 1$ for a suitable $m \in M$. But then $Z(g) \cap Z(m)$, and *a fortiori* $Z(f) \cap Z(m)$, is void; by Lemma 18, $1 \in M$.

Lemma 21 carries with it the evident corollary that the maximal right ideals in $C(X, A)$ are two-sided and the primitive ideals maximal. However in this connection it may be of interest to show that not all right ideals are necessarily two-sided.

Example. Let X be the unit interval, A the quaternions. Consider the principal right ideal generated by $f = x + ixy$, where we have written y for $\sin(1/x)$. If this ideal is two-sided we must have $jf = fg$ for some continuous g . But for $x \neq 0$ we have

$$g = f^{-1}jf = [(1 - y^2)j - 2yk]/(1 + y^2),$$

and this cannot be extended to be continuous at $x = 0$.

Let M be a maximal ideal in $C = C(X, A)$. If M is fixed, $C - M$ is isomorphic to A . If M is free, $C - M$ is still a division ring containing A — for the constant functions must be sent into separate residue classes. The $*$ -function can be extended to $C - M$, granted that $0^* = 0$; for then by Lemma 21, $f \in M$ implies $f^* \in M$, so that the $*$ -function is uniquely definable on residue classes, and $ff^* + gg^* \in M$ implies $f, g \in M$, so that the fundamental property persists.

We shall now give an explicit instance where $C - M$ is a proper extension of A . Let X be the same abstract space as A , and let f be the function which maps X identically on A . It is impossible that $f - c \in M$ for c a constant function, for $f - c$ has only one zero and by Lemma 19 this would be a zero of all of M , making M a fixed ideal. Thus f does not map into A , and $C - M$ must be a proper extension of A .

11. Functions vanishing at a point. Let $C(X, A, x)$ be the subring of $C(X, A)$ consisting of all functions vanishing at the point $x \in X$. If X is compact we may alternatively describe the functions as vanishing "at infinity" on the locally compact space $X - x$. Rings of this kind occur in the theory of normed rings and Boolean rings without unit.

THEOREM 34. *If X is a compact Hausdorff space and either (1) A is the field of reals or complexes, or (2) X is totally disconnected and A is a finite field or a field of characteristic zero, then a maximal ideal M in $C(X, A, x)$ is fixed.¹⁸*

¹⁸ We mean here that all the functions of M vanish at some point other than x .

Proof. First we prove that M is a prime ideal. The alternative is for $C - M = C(X, A, x) - M$ to be a zero ring multiplicatively and a group of p elements additively. But then if A has characteristic zero, any $f \in C$ is of the form $f = pg$, so $f \in M$ and $M = C$. If A is a finite field we have $f^{p^n} = f$ and again $M = C$.¹⁴

Suppose that M is a free ideal. Then for any $y \in X$, $y \neq x$, there is a function in M not vanishing at y . Following the method of Theorem 24 or Lemma 18 as the case may be, we can show that for any neighborhood of x , M contains a function not vanishing in its complement.

Next we show that M contains any function f vanishing in a neighborhood U of x . Let $g \in M$ be a function not vanishing in the complement V of U . Define $h = g^{-1}f$ in V , $h = 0$ in U . Then h is readily seen to be continuous, and $f = gh$ is in M .

Now suppose h is a function not contained in M . Since M is prime, also $h^2 \notin M$. Hence $(h^2, M) = C$, and $h = kh^2 + m$ for suitable $k \in C$, $m \in M$. Take a neighborhood U of x such that $hk \neq 1$ in the closure U' of U . Then $m/h = 1 - kh$ is a continuous function different from zero in U' , and so is its reciprocal. This function h/m may be extended to a continuous function r defined everywhere in X ; in case (1) this follows from Tietze's Theorem, in case (2) we may take U to be open and closed and set $r = 0$ outside U . The function hr is in C since it vanishes at x , and so hrm is in M . But $hrm - h^2$ vanishes in U and, by our above result, it is also in M . Hence $h^2 \in M$, a contradiction.

12. The ring of bounded functions. Let $C'(X, A)$ denote the ring of continuous bounded functions from X to A — boundedness being defined as in 4. The following result is easy to verify.

THEOREM 35. $C'(X, A, u)$ is a Q -ring if A is a locally bounded Q -ring; if further A is complete, so also is $C'(X, A, u)$.

It is to be observed that $C'(X, A, k)$ is not in general complete. In fact, if A is the reals and X is normal, then $C'(X, A, k)$ is dense in $C(X, A, u)$.

Suppose that A is a division ring. The spaces of maximal ideals in $C(X, A)$ and $C'(X, A)$ both provide compactifications of X . Gelfand and Kolmogoroff [8] have shown for the case where A is the reals that these compactifications are homeomorphic. Their technique does not seem to be applicable in more general cases. By a different method we are able to prove the following theorem.

¹⁴ The hypotheses on A in case (2) are only needed at this point to prove that M is prime. Still another adequate hypothesis is that A has a valuation.

THEOREM 36. *Let X be discrete and A a locally bounded division ring with the property that for any neighborhood U of 0 the set of inverses of the complement of U is bounded.¹⁵ Then the spaces of maximal ideals in $C(X, A)$ and $C'(X, A)$ are homeomorphic.*

Proof. Let M be a maximal ideal in C . $N = M \wedge C'$, being the contraction of a prime ideal, is prime, but in general is not a maximal ideal in C' . However it expands uniquely to a maximal ideal M' . To prove this we need only show that (N, f, g) is a proper ideal whenever (N, f) and (N, g) are both proper. Suppose on the contrary that $n + f + g = 1$, $n \in N$. Let U be a neighborhood of 0 in A such that $(1 + U)^{-1}$ is bounded (such a neighborhood exists because of local boundedness). Let Y be the intersection of $Z(n)$ —the set of zeros of n —and the set where f takes values in U . Let p be the function equal to 0 on Y and 1 elsewhere, q the function equal to 0 on $Z(n) - Y$ and 1 elsewhere. On the set where p vanishes, g takes values in $1 + U$ and has a bounded inverse. Hence $(g, p) = C'$. On the set where q vanishes, f takes values in the complement of U and by our assumption has a bounded inverse. Hence $(f, q) = C'$. But $pq \in N$ so that either p or q must be in N . In either case we have a contradiction.

Conversely let us start with a maximal ideal M' in C' . For any f in C , define f' to be the function with the same zeros as f , but equal to 1 elsewhere, i. e., f' is the characteristic function of the complement of $Z(f)$. We define M as follows: $f \in M$ if and only if $f' \in M'$. It is readily verified that M is a proper ideal in C . Suppose that M is not maximal, so that (M, g) is proper, $g \notin M$. If g' is the characteristic function of $Z(g)$ we have $g'g'' = 0 \in M'$ but $g' \notin M'$, hence $g'' \in M'$, a contradiction since $(g, g'') = C$.

The double correspondence thus established is in fact one-one. Starting with M , we have $f' \in M \wedge C' \subset M'$ for any $f \in M$. Hence at least all of M , and therefore precisely M , returns on applying the second correspondence. Conversely, when we begin with M' we evidently have $M \wedge C' \subset M'$ for the corresponding M . Since the expansion of $M \wedge C'$ is unique, the second correspondence must return us to M' .

Finally to prove the correspondence a homeomorphism, it will suffice (since both spaces are compact) to prove that $M' \supset \wedge M'_i$ implies $M \supset \wedge M_i$. Suppose $f \in M_i$ for all i , then $f' \in M'_i \subset M'$, whence $f \in M$.

We remark that (for the case X discrete) the space of maximal ideals

¹⁵ The field of rationals under the p -adic topology is locally bounded but does not satisfy the latter condition.

in $C(X, A)$ is the same for any division ring A , and is in fact the Čech compactification βX .

Another interesting question is the following: if M' is a free maximal ideal in C' , what is $C' - M'$? When A is the reals, complexes, or quaternions, or a finite field, it is known that $C' - M' = A$. Apparently it will require new methods to extend this result to more general cases.

UNIVERSITY OF CHICAGO.

BIBLIOGRAPHY.

1. W. Ambrose, "Structure theorems for a special class of Banach algebras," *Transactions of the American Mathematical Society*, vol. 57 (1945), pp. 364-386.
2. G. Birkhoff, "Lattice Theory," *Colloquium publications*, vol. 25 (New York, 1940).
3. I. S. Cohen, "On the structure and ideal theory of complete local rings," *Transactions of the American Mathematical Society*, vol. 59 (1946), pp. 54-106.
4. D. van Dantzig, "Zur topologische Algebra. II," *Compositio Mathematica*, vol. 2 (1935), pp. 201-223.
5. M. Deuring, "Algebren," *Ergebnisse der Mathematik*, vol. 4 (1935).
6. S. Eilenberg and S. MacLane, "Group extensions and homology," *Annals of Mathematics*, vol. 43 (1942), pp. 757-831.
7. I. Gelfand, "Normierte Ringe," *Math. Sbornik*, vol. 9 (1941), pp. 1-23.
8. I. Gelfand and A. Kolmogoroff, "On rings of continuous functions on topological spaces," *Comptes Rendus URSS (Doklady)*, vol. 22 (1939), pp. 11-15.
9. N. Jacobson and O. Taussky, "Locally compact rings," *Proceedings of the National Academy of Sciences*, vol. 21 (1935), pp. 106-108.
10. N. Jacobson, "Totally disconnected locally compact rings," *American Journal of Mathematics*, vol. 58 (1936), pp. 433-449.
11. N. Jacobson, "The radical and semi-simplicity for arbitrary rings," *American Journal of Mathematics*, vol. 67 (1945), pp. 300-320.
12. N. Jacobson, "A topology for the set of primitive ideals in an arbitrary ring," *Proceedings of the National Academy of Sciences*, vol. 31 (1945), pp. 333-338.
13. E. van Kampen, "Locally bicomact abelian groups and their character groups," *Annals of Mathematics*, vol. 36 (1935), pp. 448-463.
14. Y. Otobe, "On quasi-evaluations of compact rings," *Proceedings of the Imperial Academy of Tokyo*, vol. 20 (1944), pp. 278-282.
15. Y. Otobe, "On locally compact fields," *Japanese Journal of Mathematics*, vol. 19 (1945), pp. 189-202.
16. I. Shafarevich, "On the normalizability of topological fields," *Comptes Rendus URSS (Doklady)*, vol. 40 (1943), pp. 133-135.
17. M. H. Stone, "Applications of the theory of Boolean rings to general topology," *Transactions of the American Mathematical Society*, vol. 41 (1937), pp. 375-481.
18. A. Weil, "L'intégration dans les Groupes Topologiques et ses Applications," *Actualités Scientifiques et Industrielles* 869, Paris, 1938.

UNIQUENESS OF SOLUTIONS OF ULTRAHYPERBOLIC PARTIAL DIFFERENTIAL EQUATIONS.*

By O. G. OWENS.

Introduction. A solution in a bounded region V of the elliptic partial differential equation, $L[u] \equiv \sum_{i=1}^n u_{x_i x_i} = 0$, is unique if on the boundary of V the solution assumes prescribed values. The theorem is proved by transforming by means of Green's formula the integral

$$\int_V u L[u] dx.$$

A solution in a bounded region V of the normal hyperbolic partial differential equation, $L[u] \equiv u_{tt} - \sum_{i=1}^n u_{x_i x_i} = 0$, is unique if V is bounded by a characteristic cone of the equation and by a plane $t = \text{constant}$, and if on the plane boundary of V the solution and its normal derivative assume prescribed values. The theorem is proved by transforming the integral

$$\int_V 2u_t L[u] dx dt. \quad [1]$$

The purpose of this paper is to establish an analogue of the above classical integral uniqueness procedures for the ultrahyperbolic partial differential equation,

$$(1) \quad L[u] \equiv \sum_{i=1}^n u_{x_i x_i} - \sum_{k=1}^m u_{y_k y_k} = 0 \quad (n \geq 2, m \geq 2),$$

The integral yielding the desired uniqueness conditions for solutions of (1) is:

$$(2) \quad \int_V [2 \sum_{i=1}^n x_i u_{x_i} + nu] L[u] dx dy.$$

The resulting uniqueness theorem is applicable to a limit problem which is

* Received April 2, 1945; Revised July 20, 1946.

not a Cauchy problem [2] but a mixed problem having an elliptic and hyperbolic nature.

Uniqueness Theorems. Three typical uniqueness theorems applying to solutions of (1) are:

THEOREM 1. *A solution of (1) existing in a hyper-sphere V is unique if the solution assumes prescribed values on the surface V^* of V and if its normal derivative assumes prescribed values on one of the two regions subdividing V^* and having as boundary the intersection of V^* and the characteristic cone of (1) with vertex at the center of V .*

THEOREM 2. *A solution of (1) existing in a hyper-parallelepiped with faces parallel to the co-ordinate planes is unique if the solution assumes prescribed values on the faces and if its normal derivative is prescribed on a single face.*

THEOREM 3. *A solution of (1) existing in a bounded region C is unique if the solution assumes prescribed values on the frontier of C and if C is the region defined by slicing, with planes: $x_l = 0$, $x_l + \sum_{k=1}^m b_k y_k = \text{const.}$ ($\sum_{k=1}^m b_k^2 \geq 1$), any hyper-cylinder having generators parallel to the x_l -axis.*

Theorem 1 will now be proved. If there are two solutions of (1) existing in V and assuming the same mixed initial data on V^* , then their difference is a solution assuming vanishing data on V^* . Hence it suffices to show that the latter solution vanishes identically in V . The proof of this fact is commenced by introducing the symbol

$$\delta_{il} = \begin{cases} 1 & i \neq l \\ -1 & i = l \end{cases}$$

and forming the identity:

$$\begin{aligned} (3) \quad 2x_l u_{x_l} L[u] &= \sum_{i=1}^n \delta_{il} u_{x_i}^2 - \sum_{k=1}^m u_{y_k}^2 \\ &+ \sum_{i=1}^n \left\{ \frac{\partial}{\partial x_i} (2x_l u_{x_i} u_{x_i}) - \frac{\partial}{\partial x_l} (x_l u_{x_i}^2) \right\} \\ &- \sum_{k=1}^m \left\{ \frac{\partial}{\partial y_k} (2x_l u_{x_l} u_{y_k}) - \frac{\partial}{\partial x_l} (x_l u_{y_k}^2) \right\}. \end{aligned}$$

Now denote an element of the surface V^* by dV^* and denote the direction cosines of the outwardly directed normal at dV^* by

$$\frac{\partial x_i}{\partial \nu}, \frac{\partial y_k}{\partial \nu}.$$

Thus, integrating (3) over V ,

$$\begin{aligned} (4) \quad \int_V 2x_l u_{x_l} L[u] dx dy &= \int_V \sum_{i=1}^n \delta_{il} u_{x_i}^2 dx dy - \int_V \sum_{k=1}^m u_{y_k}^2 dx dy \\ &+ \int_{V^*} \sum_{i=1}^n \left\{ 2x_l \frac{\partial x_i}{\partial \nu} u_{x_i} u_{x_l} - x_l \frac{\partial x_l}{\partial \nu} u_{x_i}^2 \right\} dV^* \\ &- \int_{V^*} \sum_{k=1}^m \left\{ 2x_l \frac{\partial y_k}{\partial \nu} u_{x_l} u_{y_k} - x_l \frac{\partial x_l}{\partial \nu} u_{y_k}^2 \right\} dV^*. \end{aligned}$$

Furthermore, assuming the center of the sphere V to be the origin of co-ordinates and the radius of V to be one,

$$\frac{\partial x_i}{\partial \nu} = x_i, \quad \frac{\partial y_k}{\partial \nu} = y_k.$$

Hence, upon substituting these values in (4),

$$\begin{aligned} (5) \quad \int_V 2x_l u_{x_l} L[u] dx dy &= \int_V \sum_{i=1}^n \delta_{il} u_{x_i}^2 dx dy - \int_V \sum_{k=1}^m u_{y_k}^2 dx dy \\ &- \int_{V^*} \sum_{i=1}^n [x_l u_{x_i} - x_i u_{x_l}]^2 dV^* + \int_{V^*} \sum_{i=1}^n x_i^2 u_{x_i}^2 dV^* \\ &+ \int_{V^*} \sum_{k=1}^m [x_l u_{y_k} - y_k u_{x_l}]^2 dV^* - \int_{V^*} \sum_{k=1}^m y_k^2 u_{x_l}^2 dV^*. \end{aligned}$$

Therefore, summing (5) with respect to l ,

$$\begin{aligned} (6) \quad \int_V \sum_{l=1}^n 2x_l u_{x_l} L[u] dx dy &= (n-1) \int_V \sum_{i=1}^n u_{x_i}^2 dx dy - n \int_V \sum_{k=1}^m u_{y_k}^2 dx dy \\ &+ \sum_{l=1}^n \sum_{k=1}^m \int_{V^*} [x_l u_{y_k} - y_k u_{x_l}]^2 dV^* - \sum_{l=1}^n \sum_{i=1}^n \int_{V^*} [x_l u_{x_i} - x_i u_{x_l}]^2 dV^* \\ &+ \int_{V^*} \left[\sum_{i=1}^n x_i^2 - \sum_{k=1}^m y_k^2 \right] \left[\sum_{l=1}^n u_{x_l}^2 \right] dV^*. \end{aligned}$$

In addition to relation (6) we have the following:

$$(7) \quad \int_V uL[u]dxdy = \int_V \sum_{k=1}^m u_{y_k}^2 dxdy - \int_V \sum_{i=1}^n u_{x_i}^2 dxdy \\ + \int_{V^*} u \left[\sum_{i=1}^n x_i u_{x_i} - \sum_{k=1}^m y_k u_{y_k} \right] dV^*.$$

Consequently, because of (6) and (7),

$$(8) \quad \int_V \left[\sum_{i=1}^n 2x_i u_{x_i} + nu \right] L[u]dxdy = - \int_V \sum_{i=1}^n u_{x_i}^2 dxdy \\ + \sum_{i=1}^n \sum_{k=1}^m \int_{V^*} [x_i u_{y_k} - y_k u_{x_i}]^2 dV^* - \sum_{i=1}^n \sum_{i=1}^n \int_{V^*} [x_i u_{x_i} - x_i u_{x_i}]^2 dV^* \\ + n \int_{V^*} u \left[\sum_{i=1}^n x_i u_{x_i} - \sum_{k=1}^m y_k u_{y_k} \right] dV^* + \int_{V^*} \left[\sum_{i=1}^n x_i^2 - \sum_{k=1}^m y_k^2 \right] \left[\sum_{i=1}^n u_{x_i}^2 \right] dV^*.$$

Thus, recalling the assumption that u is zero on V^* and noticing that $[x_i u_{y_k} - y_k u_{x_i}]$, $[x_i u_{x_i} - x_i u_{x_i}]$ are inner derivatives on V^* ,

$$(9) \quad \int_V \left[\sum_{i=1}^n 2x_i u_{x_i} + nu \right] L[u]dxdy = - \int_V \sum_{i=1}^n u_{x_i}^2 dxdy \\ + \int_{V^*} \left[\sum_{i=1}^n x_i^2 - \sum_{k=1}^m y_k^2 \right] \left[\sum_{i=1}^n u_{x_i}^2 \right] dV^*.$$

Now, as in the case of the normal hyperbolic partial differential equation, the prescribing of limit values for a solution of (1) is influenced by its characteristic cone:

$$(10) \quad \sum_{i=1}^n x_i^2 - \sum_{k=1}^m y_k^2 = 0.$$

The effect in our case is that the normal derivative of the solution need be prescribed, as being zero, only on that portion of V^* for which

$$\sum_{i=1}^n x_i^2 - \sum_{k=1}^m y_k^2 \geq 0.$$

Since this suffices for us to infer from (9) that

$$(11) \quad \int_V \sum_{i=1}^n u_{x_i}^2 dxdy = 0.$$

Hence, since (11) implies that u does not depend on the variables x_i , u is a solution of the equation,

$$(12) \quad \sum_{k=1}^m u_{y_k y_k} = 0.$$

But (12) together with the fact that $u = 0$ on V^* assures the vanishing of u on V . Thus, Theorem 1 has been proved.

The proof of Theorem 2 will now be indicated. Firstly, it is no specialization to assume the parallelepiped V to be defined by the inequalities,

$$0 \leq x_i \leq a_i, \quad 0 \leq y_k \leq b_k \quad (a_i > 0, \quad b_k > 0),$$

and the normal derivative of the solution to be zero on the face $x_i = a_i$. Secondly, because of (5) and (7),

$$\int_V [2x_i u_{x_i} + u] L[u] dx dy = - \int_V u_{x_i}^2 dx dy,$$

granted that u vanishes on the faces of V and the normal derivative of u vanishes on the face $x_i = a_i$. Thus, since $L[u] \equiv 0$, the solution is independent of the variable x_i . Hence, as $u \equiv 0$ on the face $x_i = a_i$, $u \equiv 0$ on V .

The proof of Theorem 3 is carried through in an analogous manner.

UNIVERSITY OF CALIFORNIA,
BERKELEY, CALIFORNIA.

UNIVERSITY OF NEVADA,
RENO, NEVADA.

REFERENCES.

-
- [1] K. Friedrichs and H. Lewy, *Mathematische Annalen*, vol. 98 (1927-28), pp. 192-204.
 - [2] F. John, *Proceedings of the National Academy of Sciences*, vol. 29 (1943), pp. 98-104.

NOTE ON SOME STRONG LAWS OF LARGE NUMBERS.*

By KAI-LAI CHUNG.

Let $\{X_n\}$ denote a sequence of independent random variables with the distribution functions $\{V_n(x)\}$ and let $S_n = \sum_{v=1}^n X_v$. Let $\{a_v\}$ denote a positive increasing sequence tending to infinity.

Recently Feller¹ has proved some theorems which can be looked upon as being extensions of a theorem of Marcinkiewicz and Zygmund.² In the spirit of this extension, we shall give an analogous extension of another theorem of the latter authors.³ together with some colloraries dealing with the strong law of large numbers.

For the sake of completeness we state the main result of Feller's paper in a version suitable for our purposes. It reads as follows:

THEOREM 1 (Feller). Let $V_n(x) = V(x)$. Suppose that

$$\sum_{v=i}^{\infty} 1/a_v^2 = O(1/a_i^2)$$

and that either

$$a_n/n \uparrow$$

or

$$a_n/n \downarrow \quad \text{and} \quad \int_{-\infty}^{\infty} x dV(x) = 0.$$

Let $\psi(x)$ be a positive, even function such that $\psi(x) \uparrow$ for $x > 0$ and let

$$\psi(a_v) = v;$$

then

$$\Pr\left(\sum_{v=1}^{\infty} X_v/a_v \text{ converges}\right) = \begin{cases} 0 \\ 1 \end{cases}$$

according as

$$\int_{-\infty}^{\infty} \psi(x) dV(x) \begin{cases} = \\ < \end{cases} \infty.$$

* Received August 7, 1946.

¹ Feller, W., "A limit theorem for random variables with infinite moments," *American Journal of Mathematics*, vol. 68 (1946), pp. 257-262.

² Marcinkiewicz, J. and Zygmund, A., "Sur les fonctions indépendantes," *Fundamenta Mathematicae*, vol. 29 (1937), pp. 60-90. See Theorem 5.

³ *Loc. cit.*, see Theorem 9.

A close analogue of Theorem 1 for the more general case of unequal distributions is the following.

THEOREM 2. *Let $\psi(x)$ be a positive, even function non-decreasing for $x > 0$. Suppose that either*

$$(i) \quad \psi(x)/x \downarrow$$

or

$$(ii) \quad \psi(x)/x \uparrow \quad \psi(x)/x^2 \downarrow \text{ and } \int_{-\infty}^{\infty} x dV_v(x) = 0 \text{ for all } v.$$

Let

$$(1) \quad \int_{-\infty}^{\infty} \psi(x) dV_v(x) = M_v^{(\psi)}.$$

If the series

$$(2) \quad \sum_{v=1}^{\infty} M_v^{(\psi)}/\psi(a_v)$$

is convergent, then

$$(3) \quad \Pr\left(\sum_{v=1}^{\infty} X_v/a_v \text{ converges}\right) = 1.$$

Conversely, if the series (2) is divergent, then there exists a sequence of independent random variables $\{X_v\}$ having distribution functions $\{V_v(x)\}$ such that (1) is satisfied and

$$(4) \quad \Pr\left(\sum_{v=1}^{\infty} X_v/a_v \text{ diverges}\right) = 1.$$

Proof. Let $E(X)$ denote the mathematical expectation of X . Define

$$Y_v = \begin{cases} X_v & \text{if } |X_v| < a_v, \\ 0 & \text{if } |X_v| \geq a_v. \end{cases}$$

Consider the case (i) first. We have

$$\begin{aligned} \sum_{v=1}^{\infty} E(Y_v^2/a_v^2) &= \sum_{v=1}^{\infty} \int_{|x| < a_v} (x^2/a_v^2) dV_v(x) \\ &\leq \sum_{v=1}^{\infty} \int_{|x| < a_v} (\psi(x)/\psi(a_v)) dV_v(x) \leq \sum_{v=1}^{\infty} M_v^{(\psi)}/\psi(a_v) \end{aligned}$$

since $\psi(a_v)/a_v^2 \leq \psi(x)/x^2$ for $|x| < a_v$ by assumption (i).

Next we have

$$\begin{aligned} \sum_{\nu=1}^{\infty} |E(Y_{\nu})|/a_{\nu} &\leq \sum_{\nu=1}^{\infty} \int_{|x| < a_{\nu}} (|x|/a_{\nu}) dV_{\nu}(x) \\ &\leq \sum_{\nu=1}^{\infty} \int_{|x| < a_{\nu}} (\psi(x)/\psi(a_{\nu})) dV_{\nu}(x) \leq \sum_{\nu=1}^{\infty} M_{\nu}^{(\psi)}/\psi(a_{\nu}). \end{aligned}$$

Finally,

$$\begin{aligned} \sum_{\nu=1}^{\infty} \Pr(Y_{\nu} \neq X_{\nu}) &= \sum_{\nu=1}^{\infty} \int_{|x| \geq a_{\nu}} dV_{\nu}(x) \\ &\leq \sum_{\nu=1}^{\infty} \int_{|x| \geq a_{\nu}} (\psi(x)/\psi(a_{\nu})) dV_{\nu}(x) \leq \sum_{\nu=1}^{\infty} M_{\nu}^{(\psi)}/\psi(a_{\nu}). \end{aligned}$$

Hence (3) follows from Kolmogoroff's three-series theorem.⁴

In case (ii) the only modification is that we have

$$\begin{aligned} \sum_{\nu=1}^{\infty} |E(Y_{\nu})|/a_{\nu} &= \sum_{\nu=1}^{\infty} 1/a_{\nu} \left| \int_{|x| \geq a_{\nu}} x dV_{\nu}(x) \right| \\ &\leq \sum_{\nu=1}^{\infty} \int_{|x| \geq a_{\nu}} (|x|/a_{\nu}) dV_{\nu}(x) \leq \sum_{\nu=1}^{\infty} 1/\psi(a_{\nu}) \int_{|x| \geq a_{\nu}} \psi(x) dV_{\nu}(x) \end{aligned}$$

since now $\psi(a_{\nu})/a_{\nu} \leq \psi(x)/x$ for $|x| \geq a_{\nu}$ by assumption (ii).

Finally, if (2) is divergent, we proceed, following Kolmogoroff, by defining a sequence of independent $\{X_{\nu}\}$ as follows:

If $M_{\nu}^{(\psi)}/\psi(a_{\nu}) \leq 1$,

$$X_{\nu} = \begin{cases} \pm a_{\nu} & \text{each with probability } M_{\nu}^{(\psi)}/2\psi(a_{\nu}); \\ 0 & \text{with probability } 1 - M_{\nu}^{(\psi)}/\psi(a_{\nu}). \end{cases}$$

If $M_{\nu}^{(\psi)}/\psi(a_{\nu}) > 1$,

$$X_{\nu} = \pm \psi^{-1}(M_{\nu}^{(\psi)}) \text{ each with probability } 1/2,$$

where $\psi^{-1}(x)$ denotes the inverse function of ψ . In either case we have

$$\Pr(|X_{\nu}| \geq a_{\nu}) \geq \min(M_{\nu}^{(\psi)}/\psi(a_{\nu}), 1).$$

Hence by Borel's lemma

$$\Pr(\overline{\lim}_{\nu \rightarrow \infty} |X_{\nu}|/a_{\nu} \geq 1) = 1,$$

and (4) follows.

⁴ Kolmogoroff, A., "Über die Summen durch den Zufall bestimmten unabhängigen Grössen," *Mathematische Annalen*, vol. 99 (1928), pp. 309-319, and vol. 102 (1929), pp. 484-488.

By the well-known Kronecker theorem, we can deduce from (3) sufficient conditions for the strong law of large numbers. It is then seen that our criterion includes those of Kolmogoroff, Marcinkiewicz and Zygmund. Among the consequences, which seem to be new as far as we are aware, we point out the following cases.

COROLLARY 1. Suppose $\int_{-\infty}^{\infty} x dV_{\nu}(x) = 0$. Let $\phi(x)$ be any positive, even function non-decreasing for $x > 0$ such that

$$\sum_{\nu=1}^{\infty} 1/\nu\phi(\nu) < \infty.$$

In particular we may take $\phi(x) = |\log |x||^{1+\epsilon}$, $\epsilon > 0$. Then if

$$\int_{-\infty}^{\infty} |x| \phi(x) dV_{\nu}(x) \leq C$$

where C is independent of ν , we have

$$\Pr(\lim_{n \rightarrow \infty} s_n/n = 0) = 1.$$

COROLLARY 2. Suppose $\int_{-\infty}^{\infty} x dV_{\nu}(x) = 0$. If $1 < p \leq 2$ and for an $\epsilon > 0$

$$\int_{-\infty}^{\infty} |x|^p |\log |x||^{1+\epsilon} dV_{\nu}(x) \leq C$$

where C is independent of ν , we have

$$\Pr(\lim_{n \rightarrow \infty} s_n/n^{1/p} = 0) = 1.$$

These results may be compared with those for the case of equal distributions, where the conditions are weaker in that the logarithm factor in the above is omitted. Our results however are the best possible in the sense of the second part of Theorem 2.

PRINCETON UNIVERSITY,
PRINCETON, N. J.

ON THE ERGODIC THEOREMS.*

By PHILIP HARTMAN.

It was recently suggested to me by Professor Wintner that it should be possible to find a proof of the Birkhoff ergodic theorem [1] along the lines of the simplest proof of the fundamental theorem of calculus, namely, a proof depending on a lemma of F. Riesz [4]. The role of Riesz's lemma in the Lebesgue theory of differentiation is the avoidance of Vitali's covering theorem on which the earlier proofs of the fundamental theorem of calculus were based. In the proof of the ergodic theorem, as given by Birkhoff and his successors, the corresponding role of Vitali's covering theorem is represented by Birkhoff's decomposition of the image sets or by other covering theorems. As will be seen in the proof of (I), below, an appeal to the Riesz lemma leads easily to an improved form of the fundamental inequality on which Birkhoff's proof of the ergodic theorem depends. This improved inequality was also obtained by Kakutani and Yosida [3] using the methods of Birkhoff.

For the sake of completeness, the Birkhoff ergodic theorem will be deduced from this inequality in (II) and the (L^p) , $p \geq 1$, mean ergodic theorem will then be derived from (I) and (II). This procedure is precisely the reverse of that applied by Wiener [5] and his followers.

For earlier references, see Hopf [2].

In what follows, Ω denotes a space of points P carrying a non-negative measure μ , for which $\mu(\Omega) = 1$. Let $P_t = \tau_t P$, where $-\infty < t < \infty$, be a group of measure-preserving transformations of Ω into itself; by this is meant that

$$(1) \quad \tau_u(\tau_v P) = \tau_{u+v} P$$

and each transformation τ_t sends measurable sets into measurable sets of the same measure. Let $f(P)$ be a real-valued function of class $(L^1) = (L)$ on Ω , that is, $f(P)$ is measurable and

$$(2) \quad \int_{\Omega} |f(P)| d\mu < \infty.$$

In addition, it will be supposed that $f(\tau_t P)$ is a measurable function on the

* Received February 19, 1947.

product space of Ω and the line $-\infty < t < \infty$, where measure on the line is the ordinary Lebesgue measure. If these assumptions are satisfied and if a zero set of points P is excluded, the integral

$$\int_0^t f(\tau_u P) du$$

exists for all finite values of t by virtue of the measure-preserving character of τ_t and by Fubini's theorem on iterated integrals.

A subset S of Ω will be said to be invariant if it is measurable and if its image $\tau_t S$, under each transformation τ_t , differs from S only by a zero set, which may depend on t .

(I) Let α be an arbitrary number and let U denote the set of points P of Ω for which

$$\text{l. u. b. } t^{-1} \int_0^t f(\tau_u P) du > \alpha.$$

Then, if S is any invariant subset of Ω ,

$$(3) \quad \alpha \mu(S \cdot U) \leq \int_{S \cdot U} f(P) d\mu.$$

In particular, if S is contained in U ,

$$(3 \text{ bis}) \quad \alpha \mu(S) \leq \int_S f(P) d\mu.$$

The form of Riesz's lemma [4] which will be needed for the proof of (I) is as follows:

Let $g(x)$ be a real-valued continuous function on an interval $a \leq x \leq b$ and let α be any fixed number. Let E denote the set of interior points x of (a, b) for which there exists at least one number x' satisfying $a \leq x' < x$ and $g(x) - g(x') > \alpha(x - x')$. Then E is an open set and the inequality $\alpha(b_n - a_n) \leq g(b_n) - g(a_n)$ holds for every n if $\Sigma(a_n, b_n)$ is the decomposition of E into disjoint open intervals.

Proof of (I). For a fixed t , let $U(t)$ denote the set of points P of Ω for which

$$\text{l. u. b. } v^{-1} \int_0^v f(\tau_u P) du > \alpha.$$

Then $U(t)$ is a measurable set which is non-decreasing with t and its limit set is U ,

$$(4) \quad U = \lim_{t \rightarrow \infty} U(t).$$

Let $S(t)$ denote the image of $U(t)$ under the transformation τ_t , so that $S(t) = \tau_t U(t)$ or $U(t) = \tau_{-t} S(t)$. By (1),

$$(5) \quad \int_{t-v}^t f(\tau_{-u} P) du = \int_0^v f(\tau_u \tau_{-t} P) du,$$

hence $S(t)$ is the set of points P for which

$$(6) \quad \text{l. u. b. } v^{-1} \int_{t-v}^t f(\tau_{-u} P) du > \alpha.$$

For a point P and a positive number m , let $E(P, m)$ denote the set of t -values on $(0, m)$ for which (6) holds. If $g(x)$ is defined to be

$$\int_0^x f(\tau_{-u} P) du, \quad (0 \leq x \leq m),$$

then the set $E(P, m)$ has the same significance as the set E in Riesz's lemma. Hence, on adding the inequalities implied by this lemma,

$$\alpha \Sigma (b_n - a_n) \leq \Sigma (g(b_n) - g(a_n))$$

or

$$\alpha \int_E dt \leq \int_E f(\tau_{-t} P) dt; \quad (E = E(P, m)).$$

Integrating this inequality over the measurable set S , one obtains

$$\alpha \int_S \left(\int_E dt \right) d\mu \leq \int_S \left(\int_E f(\tau_{-t} P) dt \right) d\mu.$$

Hence, by Fubini's theorem on iterated integrals,

$$\alpha \int_0^m \mu(S \cdot S(t)) dt \leq \int_0^m \left(\int_{S \cdot S(t)} f(\tau_{-t} P) d\mu \right) dt.$$

Since S is an invariant set, it follows that, when t is fixed, $\tau_{-t}(S \cdot S(t)) = S \cdot \tau_{-t} S(t) = S \cdot U(t)$ if a zero set of points P of Ω is ignored. Consequently,

$$\int_{S \cdot S(t)} f(\tau_{-t} P) d\mu = \int_{S \cdot U(t)} f(P) d\mu$$

and

$$\mu(S \cdot S(t)) = \mu(S \cdot U(t))$$

since τ_{-t} is measure-preserving. Consequently, the last inequality can be written in the form

$$(7) \quad \alpha \int_0^m \mu(S \cdot U(t)) dt \leq \int_0^m \int_{S \cdot U(t)} f(P) d\mu dt.$$

On the other hand, the limit relation (4) implies that

$$\mu(S \cdot U(t)) \rightarrow \mu(S \cdot U) \text{ and } \int_{S \cdot U(t)} f(P) d\mu \rightarrow \int_{S \cdot U} f(P) d\mu \quad (t \rightarrow \infty).$$

Therefore, if the inequality (7) is divided by m , the assertion (3) follows by letting $m \rightarrow \infty$. This completes the proof of (I).

An application of (I) to the function $-f(P)$ and to the constant $\alpha = -\beta$ gives

(I bis) *Let β be an arbitrary number and let V denote the set of points P of Ω for which*

$$(8) \quad \text{g. l. b. } t^{-1} \int_0^t f(\tau_u P) du < \beta.$$

Then, if S is any invariant subset of Ω ,

$$(9) \quad \beta \mu(S \cdot V) \geq \int_{S \cdot V} f(P) d\mu;$$

and, if S is contained in V ,

$$(9 \text{ bis}) \quad \beta \mu(S) \geq \int_S f(P) d\mu.$$

The statements (I) and (I bis) imply the following theorem, which contains the Birkhoff ergodic theorem:

(II) *For almost all points P of Ω ,*

$$(10) \quad f^*(P) = \lim_{t \rightarrow \infty} t^{-1} \int_0^t f(\tau_u P) du$$

exists, is a function of class (L), and

$$(11) \quad \int_{\Omega} f^*(P) d\mu = \int_{\Omega} f(P) d\mu.$$

Proof of (II). Let α and β be any pair of numbers satisfying $\alpha > \beta$ and let S denote the set of points P for which

$$\liminf_{t \rightarrow \infty} t^{-1} \int_0^t f(\tau_u P) du < \beta < \alpha < \limsup_{t \rightarrow \infty} t^{-1} \int_0^t f(\tau_u P) du.$$

Then, in virtue of (1), S is an invariant set. Also, S is contained in the sets

U and V , defined in (I) and (I bis), respectively. Hence, (I) and (I bis) imply that

$$\alpha\mu(S) \leq \int_S f(P) d\mu \leq \beta\mu(S).$$

Since $\alpha > \beta$ and $\mu(S) \leq \mu(\Omega) = 1$, it follows that $\mu(S) = 0$. Hence, the limit (10) exists for almost all P ; where, however, the possibility that this limit is $\pm \infty$ is not yet excluded.

In order to see that $f^*(P)$ is finite for almost all P , let S denote the set of points P for which $f^*(P) = +\infty$. Then S is an invariant subset of Ω and S is contained in the set U of (I) for every constant α . Thus (3 bis) holds for every α . Hence, by (2), $\mu(S) = 0$. It is similarly shown that the set of points P for which $f^*(P) = -\infty$ is a zero set.

Finally, to prove that $f^*(P)$ is of class (L), let α and β be any pair of numbers such that $\alpha < \beta$ and let S be the set of points P for which $f^*(P)$ exists and satisfies

$$\alpha < f^*(P) < \beta.$$

Then, (I) and (I bis) imply that

$$\alpha\mu(S) \leq \int_S f(P) d\mu \leq \beta\mu(S).$$

It follows, from the definition of Lebesgue integrals in terms of upper and lower sums connected with a mesh on the f^* -axis, that $f^*(P)$ is of class (L) and that its integral satisfies (11).

This completes the proof of (II).

(III) *If the function $f(P)$ in (II) is of class (L^p) , $p \geq 1$, on Ω then so is the function $f^*(P)$, defined by (10), and*

$$(12) \quad \int_{\Omega} |f^*(P) - t^{-1} \int_0^t f(\tau_u P) du|^p d\mu \rightarrow 0 \text{ as } t \rightarrow \infty.$$

Proof of (III). By Hölder's inequality,

$$(13) \quad |t^{-1} \int_0^t f(\tau_u P) du|^p \leq t^{-1} \int_0^t |f(\tau_u P)|^p du.$$

The arguments used before the statement of (I) show that, for almost all P , the integral on the right of (13) exists for all finite t and is of class (L) for every fixed t . Hence, the integral in (12) exists if $f^*(P)$ is of class (L^p) . But, by (10) and (13),

$$|f^*(P)|^p \leq \lim_{t \rightarrow \infty} t^{-1} \int_0^t |f(\tau_u P)|^p du,$$

where the limit on the right exists for almost all P and is of class (L) by virtue of an application of (II) to the function $|f(P)|^p$. Consequently, $f^*(P)$ is of class (L^p) .

The assertion (10) of (II) is equivalent to

$$(14) \quad |f^*(P) - t^{-1} \int_0^t f(\tau_u P) du|^p \rightarrow 0, \quad (t \rightarrow \infty),$$

while the assertion of (III) is that (14) can be integrated term-by-term. On the other hand, this term-by-term integration is allowed if

$$(15) \quad \int_T |f^*(P) - t^{-1} \int_0^t f(\tau_u P) du|^p d\mu \rightarrow 0 \text{ as } \alpha \rightarrow \infty$$

holds uniformly in t , where $T = T_\alpha$ is the set of points P for which

$$(16) \quad \text{l. u. b.}_{0 < t < \infty} |f^*(P) - t^{-1} \int_0^t f(\tau_u P) du| > \alpha.$$

By the Minkowski inequality, the integral in (15) is majorized by the two integrals

$$(17) \quad \int_T |f^*(P)|^p d\mu$$

and

$$\int_T |t^{-1} \int_0^t f(\tau_u P) du|^p d\mu.$$

But the Hölder inequality (13) shows that the last integral does not exceed

$$\int_T (t^{-1} \int_0^t |f(\tau_u P)|^p du) d\mu,$$

which, by Fubini's theorem on iterated integrals, has the value

$$t^{-1} \int_0^t \left(\int_T |f(\tau_u P)|^p d\mu \right) du$$

and is, therefore, majorized by

$$(18) \quad \text{l. u. b.}_{0 < u < \infty} \int_W |f(P)|^p d\mu, \quad (W = \tau_u T).$$

In virtue of the absolute continuity of the integrals

$$\int |f^*(P)|^p d\mu \text{ and } \int |f(P)|^p d\mu,$$

and the measure-preserving character of the transformations τ_u , the quantities (17) and (18) tend to 0 as $\alpha \rightarrow \infty$ if the measure of $T = T_\alpha$ does.

By (1), the function (10) is invariant, that is $f^*(\tau_t P) = f^*(P)$ for every t . Hence,

$$|f^*(P) - t^{-1} \int_0^t f(\tau_u P) du| \leq t^{-1} \int_0^t |f^*(\tau_u P) - f(\tau_u P)| du,$$

and so the set $U = U_\alpha$ of points P for which

$$\liminf_{0 < t < \infty} t^{-1} \int_0^t |f^*(\tau_u P) - f(\tau_u P)| du > \alpha$$

contains the set $T = T_\alpha$. If, in (I), the function $f(P)$ is replaced by $|f^*(P) - f(P)|$ and the invariant set S is chosen to be the entire space Ω , (3) gives

$$\alpha_\mu(U) \leq \int_U |f^*(P) - f(P)| d\mu.$$

Thus,

$$\mu(T) \leq \mu(U) \leq \alpha^{-1} \int_\Omega |f^*(P) - f(P)| d\mu \rightarrow 0, \text{ as } \alpha \rightarrow \infty.$$

This completes the proof of (III).

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

-
- [1] G. D. Birkhoff, "Proof of the ergodic theorem," *Proceedings of the National Academy of Sciences*, vol. 18 (1931), pp. 656-660.
 - [2] E. Hopf, *Ergodentheorie*, *Ergebnisse der Mathematik und ihrer Grenzgebiete*, vol. 5 (1937), no. 2.
 - [3] S. Kakutani and K. Yosida, "Birkhoff's ergodic theorem and the maximal ergodic theorem," *Proceedings of the Imperial Academy of Tokyo*, vol. 15 (1939), pp. 165-168.
 - [4] F. Riesz, "Sur l'existence de la dérivée des fonctions monotones et sur quelques problèmes qui s'y rattachent," *Acta Litterarum ac Scientiarum* (Szeged), vol. 5 (1932), pp. 208-221.
 - [5] N. Wiener, "The ergodic theorem," *Duke Mathematical Journal*, vol. 5 (1939), pp. 1-18.

MAPPING THEOREMS FOR NON-COMPACT SPACES.*

By C. H. DOWKER.

The chief aim of this paper is to extend to more general spaces the theorems of Hopf [17]¹ and Bruschlinsky [7] on the mappings of complexes in spheres. Since these theorems can be stated most neatly in terms of co-homologies, the latter are used throughout the paper.

It is shown that, if cocycles based on infinite coverings are used, the theorems of Hopf and Bruschlinsky can be extended to all paracompact² normal spaces, a class of spaces which includes in particular the compact³ Hausdorff spaces and the metric separable spaces. With the traditional Čech cocycles based on finite coverings, the theorems hold for countably compact normal spaces. If the cocycles are based on the finite coverings of compact subsets, Bruschlinsky's theorem, but not Hopf's, extends to a certain class⁴ of non-compact spaces. Keeping cocycles based on finite coverings but replacing homotopy by uniform homotopy in the statements of Hopf's and Bruschlinsky's theorems, one can extend these theorems to arbitrary normal spaces.

We call attention to the following additional results: (a) The 1-dimensional Čech co-Betti number of the straight line is the power of the continuum. (b) The covering dimension of a normal space is the same whether based on finite coverings or on more general locally finite coverings. (c) Borsuk's theorem ([5], p. 103) on extensions of uniformly homotopic mappings into absolute neighborhood retracts holds for normal spaces, and the corresponding theorem with ordinary homotopy holds for paracompact normal spaces.

1. Coverings and nerves. By a covering of a space⁵ we mean a

* Received May 31, 1940; Revised October 24, 1946. The author wishes to thank E. G. Begle for his assistance in an extensive revision of this paper. Some of the results in the paper were reported in the *Proceedings of the National Academy of Sciences*, vol. 23 (1937), pp. 293-294. Section 3 was presented to the American Mathematical Society February 23, 1946, under the title of "Lebesgue dimension of a normal space."

¹ Numbers in square brackets refer to the bibliography.

² Paracompact spaces have been investigated by J. Dieudonné [11].

³ Compact means bicomact.

⁴ This is the class of locally connected locally compact paracompact normal spaces.

⁵ In general no separation axiom is assured. Thus, using the nomenclature of Alexandroff-Hopf [3], a space is required to be a topological space but not necessarily a T_0 space.

covering by a finite or infinite collection of open sets. By the nerve of a covering we mean the nerve realized as a space with the Euclidean metric as defined by S. Lefschetz ([23], p. 5). Each point p of such a Euclidean complex has a set of barycentric coordinates $\{x_i\}$, a coordinate x_i being associated with each vertex a_i of the complex, such that $\sum x_i = 1$, and such that $0 \leq x_i \leq 1$ for every i , $x_i > 0$ if and only if a_i is a vertex of the simplex containing p , and $x_i = 1$ if and only if p is the vertex a_i . The distance between any two points $\{x_i\}$ and $\{y_i\}$ of the complex is given by the formula $(\sum |x_i - y_i|^2)^{1/2}$.

A covering of a space is called *finite* if it consists of a finite number of open sets. A space is called compact if every covering of the space has a finite refinement.⁶ Čech's homology theory [8] is based on finite coverings and has proved very useful for investigating compact spaces. For non-compact spaces, even those as simple as the straight line, this theory gives intuitively unsatisfactory results as is shown in Theorem 9.6. As might be supposed, the coverings appropriate for use with non-compact spaces are infinite. We shall see that important theorems can be extended to a broad class of non-compact spaces if the finiteness conditions are relaxed.

A covering of a space is called *locally finite*⁷ if each point of R has a neighborhood meeting only a finite number of sets of the covering. A space is called *paracompact* if every covering of the space has a locally finite refinement. Paracompact spaces are thus a generalization of compact spaces. J. Dieudonné [11] has shown that many properties of compact spaces extend to paracompact spaces. In particular, he has shown that for paracompact spaces normality is less restrictive than the Hausdorff separation axiom. In the theorems of this paper we require that paracompact spaces be normal but not necessarily Hausdorff.

A covering of a space is called *star-finite* if each set of the covering meets only a finite number of other sets of the covering. A space is called an *s-space* if every covering has a star-finite refinement. The star-finite coverings include the finite coverings and are included in the locally finite coverings. They are from one point of view the most immediate generalization of finite coverings, for the star-finite coverings have as nerves locally finite complexes, and locally finite complexes are the simplest type after the finite complexes. In the same

⁶ A covering \mathcal{U} of R is a refinement of a covering \mathcal{B} if every open set of \mathcal{U} is contained in some open set of \mathcal{B} .

⁷ Locally finite = neighborhood-finite. "Locally finite" has been used by S. Lefschetz ([24], p. 13) to mean star-finite, and by A. Weil ([26], p. 34) to mean that each compact set meets a finite number of open sets of the covering.

sense the s -spaces are the most immediate generalization of the compact spaces. The generalization is moreover non-trivial since, as has been shown by S. Kaplan [21], every separable metric space is an s -space. The paracompact spaces are a further non-trivial generalization. An infinite complex is paracompact⁸ but it need not be an s -space.

In combinatorial topology the chief interest in locally finite coverings results from the fact that the theorem of Alexandroff ([1], p. 121)⁹ on mapping a space into the nerve of a covering can be extended to normal spaces if the coverings are required to be locally finite. Theorem 1.1, which is a form of Alexandroff's theorem, is the basic lemma of this paper, being used in the proof of all the remaining theorems.

Let \mathcal{U} be any covering of a space R and let N be the nerve of \mathcal{U} . A mapping ϕ of R into N is called *canonical* relative to \mathcal{U} if the inverse image of the star of each vertex of N is contained in the open set of \mathcal{U} corresponding to this vertex. Let $\{U_i\}$, where each i is a transfinite ordinal, be the open sets of \mathcal{U} and let $\{u_i\}$ be the corresponding vertices of N . Let $St(u_i)$ be the star of u_i in N . Then ϕ is canonical when $\phi^{-1}St(u_i) \subset U_i$, for each $U_i \in \mathcal{U}$.

If \mathcal{U} is a point-finite covering, i. e., if each point of R is in only a finite number of sets of \mathcal{U} , there is another useful formulation of the condition that a mapping of R in the nerve of \mathcal{U} be canonical. Let p be any point of R . The vertices of N corresponding to those open sets of the point-finite covering \mathcal{U} which contain p are the vertices of a certain simplex $\sigma(p)$ of N . This simplex $\sigma(p)$ is called the simplex of N determined by p . A *necessary and sufficient condition* that a mapping ϕ of R in the nerve N of a point finite covering \mathcal{U} be canonical is that for every point $p \in R$, $\phi(p)$ is in the closure $\overline{\sigma(p)}$ of the simplex of N determined by p . For, if ϕ is canonical, let p be any point of R . Then, since $\phi(R) \subset N$, $\phi(p) \in \sigma$ for some simplex σ of N . Let u_i be any vertex of σ . Then $\phi(p) \in St(u_i)$, and hence $p \in \phi^{-1}St(u_i)$. Hence $p \in U_i$ and u_i is a vertex of $\sigma(p)$. Hence all vertices of σ are vertices of $\sigma(p)$ and $\sigma \subset \overline{\sigma(p)}$. Hence $\phi(p) \in \overline{\sigma(p)}$. On the other hand, let $\phi(p) \in \overline{\sigma(p)}$ for each point $p \in R$. Let U_i be an arbitrary open set of \mathcal{U} and let \underline{p} be an arbitrary point of $\phi^{-1}St(u_i)$. Then $\phi(\underline{p}) \in St(u_i)$ and also $\phi(\underline{p}) \in \overline{\sigma(p)}$. Hence u_i is a vertex of $\sigma(p)$. Hence $p \in U_i$. Hence $\phi^{-1}St(u_i) \subset U_i$. Therefore ϕ is canonical.

THEOREM 1.1. *If \mathcal{U} is a locally finite covering of a normal space R there is a canonical mapping ϕ of R into the nerve N of \mathcal{U} .*

⁸ The proof of this statement will be published in a forthcoming paper by the author.

⁹ See also [22] and [18]. The extension to star-finite coverings of normal spaces is given by S. Lefschetz ([25], p. 41).

Proof. Let $\{U_i\}$ ($i = 1, 2, \dots, \alpha$, where α is a finite or transfinite ordinal number) be the open sets of \mathbb{U} and let $\{u_i\}$ be the corresponding vertices of K . Let $F_{ijk\dots m} = R - \sum_{r \neq i, j, \dots, m} U_r$ be defined for every set of subscripts (i, j, \dots, m) such that $U_i U_j \dots U_m \neq 0$, i. e., such that $u_i u_j \dots u_m$ is a simplex of N . Any F may of course be the null set, but each point of R is contained in some F . Each F is closed.

A mapping ϕ of R into N will be defined by assigning to each point of R the barycentric coordinates of the corresponding point of N . A point of the closed simplex $u_i u_j \dots u_m$ of N has coordinates $x_i^{(ij\dots m)}$, $x_j^{(ij\dots m)}$, \dots , $x_m^{(ij\dots m)}$ where $x_i^{(ijk\dots m)} = x_i^{(jik\dots m)}$, etc. If the point is on the boundary simplex $u_j \dots u_m$, its coordinates satisfy the relations $x_j^{(ij\dots m)} = x_j^{(j\dots m)}$, etc.

Let $x_i^{(ij)} = x_i^{(ij)}(p)$ be a continuous function of p for p in F_{ij} such that $0 \leq x_i^{(ij)}(p) \leq 1$, $x_i^{(ij)}(p) = 0$ if $p \in F_j$ and $x_i^{(ij)} = 1$ if $p \in F_i$. The existence of such a function is proved by Urysohn's lemma. Let $x_j^{(ij)} = 1 - x_i^{(ij)}$. In this way we determine the barycentric coordinates of the image of each F_{ij} . We proceed by recursion. Assume that the coordinates of each $F_{ijk\dots m}$ having q subscripts have been determined. Let $F_{ijk\dots mn}$ have $q + 1$ subscripts. Let

$$\begin{aligned} y_i^{(ijk\dots mn)} &= x_i^{(ik\dots mn)} & \text{for } p \in F_{ik\dots mn}, \\ &\dots \\ &= x_i^{(ijk\dots m)} & \text{for } p \in F_{ijk\dots m}, \\ &= 0 & \text{for } p \in F_{jk\dots mn}. \end{aligned}$$

It is thus defined as a continuous function of p for p in a certain closed set of $F_{ijk\dots mn}$, and hence the function can be extended to a continuous function on the whole of $F_{ijk\dots mn}$. Similarly $y_j^{(ijk\dots mn)}$, \dots , $y_n^{(ijk\dots mn)}$ are defined as functions of p for $p \in F_{ijk\dots mn}$. Let

$$\begin{aligned} y_n^{(ijk\dots mn)} &= x_n^{(jk\dots mn)} & \text{for } p \in F_{jk\dots mn}, \\ &\text{etc.} \\ &= 0 & \text{for } p \in F_{ijk\dots m}, \\ &= 1 \end{aligned}$$

for p in the closed set where $y_i^{(ijk\dots mn)}(p) = y_j^{(ijk\dots mn)}(p) = \dots = y_m^{(ijk\dots mn)} = 0$. Let

$$x_i^{(ijk\dots mn)} = \frac{y_i^{(ij\dots n)}}{y_i^{(ij\dots n)} + y_j^{(ij\dots n)} + \dots + y_n^{(ij\dots n)}},$$

etc. Thus the mapping is defined for $F_{ijk\dots mn}$ and reduces on the "boundary" sets $F_{ijk\dots m}$ etc. to that previously defined. Thus the mapping can be extended to all sets $F_{ijk\dots mn}$ with $q+1$ subscripts. Proceeding *ad inf.*, the mapping can be extended to a mapping ϕ of the whole space R into N .

The mapping ϕ is continuous on each closed set $F_{ij\dots}$ and hence on the sum of any finite number of these. Since the covering \mathfrak{U} is locally finite each point of R has a neighborhood containing points of only a finite number of the U 's and hence contained in a finite collection of the F 's. Therefore ϕ is continuous in the neighborhood of each point of R and hence is a continuous mapping.

Let the simplex of N determined by p be $\sigma(p) = u_i u_j \cdots u_m$. Then p is not in U_r for $r \neq i, j, \dots, m$. Hence $p \in \overline{F_{ij\dots m}}$. But ϕ maps $F_{ij\dots m}$ into the closure of $u_i u_j \cdots u_m$. Hence $\phi(p) \in \overline{\sigma(p)}$. Hence ϕ is canonical. This completes the proof.

It may be noted that a simplicial mapping of one infinite complex in another is not necessarily continuous. Thus a mapping of an infinite complex K may be discontinuous even if each finite subcomplex of K is mapped continuously.

LEMMA 1.2. *Let \mathfrak{U} be a locally finite covering of a normal space R and let ϕ be a canonical mapping of R in the nerve N of \mathfrak{U} . Let N be mapped into a space S by a mapping f which is continuous on each finite subcomplex of N . Then the mapping $f\phi$ of R into S is continuous.*

Proof. Let p be any point of R . Since \mathfrak{U} is locally finite the point p has a neighborhood W which intersects only a finite number of the open sets U_i of \mathfrak{U} . The corresponding vertices u_i of N determine a finite subcomplex of N which contains $\phi(W)$. This finite subcomplex is mapped continuously by f . Thus $f\phi$ maps W continuously into S so that $f\phi$ is continuous in the neighborhood of each point p of R and hence is continuous.

2. Homotopy. Let R and S be two spaces and let I be the closed line segment $[0, 1]$. Two mappings, f_0 and f_1 , of R into S are said to be homotopic if there is a mapping h of $R \times I$ into S such that, for each $x \in R$, $h(x, 0) = f_0(x)$ and $h(x, 1) = f_1(x)$. The mapping h is called a homotopy of f_0 and f_1 . The homotopy is said to be a uniform homotopy if S is a metric space and if for every $\epsilon > 0$ there is a $\delta > 0$ such that if $|t - t'| < \delta$ then the distance $\rho(h(x, t), h(x, t')) < \epsilon$ for all x in R . In this case f_0 is said to be uniformly homotopic to f_1 .

In general the property of being a uniform homotopy depends on the metric of S . However, if S is a compact metric space the property is a topological one. For, if ϕ is a homeomorphism of S onto a space S_1 , then ϕ is uniformly continuous, and hence if $\epsilon > 0$ there is a number $\eta > 0$ such that if the distance between two points of S is less than η their images by ϕ are less than ϵ apart. Let h be a uniform homotopy of two mappings of R into S . Let $\delta > 0$ be chosen so that if $|t - t'| < \delta$ then $\rho(h(x, t), h(x, t')) < \eta$. Then $\rho(\phi h(x, t), \phi h(x, t')) < \epsilon$ and hence ϕh is a uniform homotopy.

Both homotopy and uniform homotopy are equivalence relations, so the set of all mappings of R into S can be uniquely divided into classes of homotopic or uniformly homotopic mappings.

Borsuk's theorem ([5], p. 103) on the extensions of uniformly homotopic mappings can be extended to normal spaces as follows:

THEOREM 2.1. (*Borsuk's theorem*). *Let R' be a closed subset of a normal space R and let f_0 and f_1 be two uniformly homotopic mappings of R' into an absolute neighborhood retract¹⁰ S . If f_0 can be extended to a mapping F_0 of R into S then f_1 can be similarly extended to F_1 with F_0 and F_1 uniformly homotopic.*

Proof. We assume that S is a subset of the Hilbert cube I_ω . Let η be a number such that the closed η -neighborhood \bar{U} of S , i. e., the set of all points of I_ω whose distance from S is less than or equal to η , can be retracted onto S .

It follows immediately from the definition of uniform homotopy that there is a finite number of mappings $h(x, 0), h(x, \delta), h(x, 2\delta), \dots$ beginning with f_0 and ending with f_1 such that the distance between any two successive mappings is less than η . It is therefore sufficient to prove the theorem under the assumption that $\rho(f_0(x), f_1(x)) < \eta$ for every x in R' .

Since R is normal f_1 can be extended to a mapping G of R into I_ω . Let V be the set of points $x \in R$ such that $\rho(F_0(x), G(x)) < \eta$. Then V is an open set and, since we assume $\rho(f_0(x), f_1(x)) < \eta$ for $x \in R'$, $V \supset R'$. Let $s(x)$ be a continuous function of x , for $x \in R$, such that $0 \leq s(x) \leq 1$, $s(x) = 0$ if $x \in R - V$, and $s(x) = 1$ if $x \in R'$.

Let h_1 be the mapping of $R \times I$ into I_ω which maps (x, t) in the point dividing the segment $[F_0(x), G(x)]$ in the ratio $t:1-t$. Then h_1 is a uniform homotopy. Also h_1 maps $V \times I$ into \bar{U} for, if $x \in V$, $F_0(x) \in S$ and $\rho(F_0(x), h_1(x, t)) \leq \rho(F_0(x), G(x)) < \eta$. Let h_2 be the mapping of $R \times I$

¹⁰ The definition of absolute neighborhood retract is given by K. Borsuk ([4], p. 222).

into \bar{U} defined by $h_2(x, t) = h_1(x, t \cdot s(x))$. Then h_2 is a uniform homotopy, and, if $x \in R'$, $s(x) = 1$ and $h_2(x, 1) = h_1(x, 1) = G(x) = f_1(x)$. Let ϕ be a retraction of \bar{U} onto S , and let h_3 be the mapping of $R \times I$ into S defined by $h_3(x, t) = \phi h_2(x, t)$. Since \bar{U} is a closed, and hence compact, subset of I_ω , therefore ϕ is uniformly continuous, and therefore h_3 is a uniform homotopy. If $x \in R'$, $f_1(x) \in S$ and hence $h_3(x, 1) = \phi h_2(x, 1) = \phi f_1(x) = f_1(x)$. If $x \in R$, $F_0(x) \in S$ and hence $h_3(x, 0) = \phi h_2(x, 0) = \phi h_1(x, 0) = \phi F_0(x) = F_0(x)$. Hence, if, for $x \in R$, $F_1(x) = h_3(x, 1)$, F_1 is an extension of f_1 and is uniformly homotopic to F_0 . This completes the proof.

3. Dimension. In our presentation of Hopf's theorems we use the cohomology theory based on some family of coverings, usually the family of the finite coverings, the family of the locally finite coverings, or some family of coverings cofinal with one of these. Since the covering dimension also appears in the statements of these theorems it would be convenient in each case to base both the dimension theory and the cohomology theory on the same family of coverings. On the other hand it is not desirable to assign several dimension numbers to one space. In this section we show that the dimension of a normal space is the same whether based on finite, star-finite, or locally finite coverings of the space.¹¹

Accordingly, we temporarily assign three dimension numbers to each space. Thus $\dim_F R \leq n$ means that every finite covering of R has a finite refinement of order $\leq n + 1$, $\dim_S R \leq n$ means that every star-finite covering of R has a star-finite refinement of order $\leq n + 1$, and $\dim_L R \leq n$ means that every locally finite covering of R has a locally finite refinement of order $\leq n + 1$. In Theorem 3.5 we show that if R is a normal space these three definitions of dimension are equivalent. The following four lemmas are used in the proof of this theorem.

LEMMA 3.1. *Let f be a mapping of a closed set R' of a normal space R in the n -sphere S^n . Let $\bar{\sigma}^n$ be a closed n -simplex of a subdivision of S^n and let S^{n-1} be its boundary. If the partial¹² mapping $f|_{f^{-1}S^{n-1}}$ can be extended to a mapping g of $f^{-1}\bar{\sigma}^n$ into S^{n-1} , then f can be extended to a mapping of R into S^n .*

¹¹ An example considered by J. Dieudonné ([11], p. 67) shows that a normal space can have zero dimension in terms of finite or locally finite coverings but infinite dimension in terms of arbitrary coverings.

¹² If f is a mapping defined on a set A and if B is a subset of A we denote by $f|_B$ the mapping f restricted to B .

Proof. Since any two mappings of a space into a closed n -cell are uniformly homotopic the mapping g of $f^{-1}\bar{\sigma}^n$ into $\bar{\sigma}^n$ is uniformly homotopic to the partial mapping $f|f^{-1}\bar{\sigma}^n$. Therefore, if f_1 is the mapping of R' in $S^n - \sigma^n$ which coincides with g on $f^{-1}\bar{\sigma}^n$ and coincides with f on $f^{-1}(S^n - \sigma^n)$, then f_1 and f are uniformly homotopic mappings of R' into S^n . But, by Tietze's Lemma, the mapping f_1 of R' into the n -cell $S^n - \sigma^n$ can be extended to a mapping of R in $S^n - \sigma^n \subset S^n$. Therefore, by Theorem 2.1, f can be extended to a mapping of R in S^n .

LEMMA 3.2. *If N is an arbitrary complex with the natural metric, the covering of N by the stars of its vertices has a refinement \mathfrak{B} such that, in the nerve of the covering \mathfrak{B} , the star of each vertex is finite dimensional.*

Proof. Let $\{u_i\}$ be the vertices of N and let $St(u_i)$ be the star of u_i . Let B_i be the boundary of the star of u_i ; $B_i = \overline{St(u_i)} - St(u_i)$. Let N^n be the n -dimensional part of N , i. e., the complex consisting of the simplexes of dimension $\leq n$.

We define a sequence of neighborhoods of the simplexes of N as follows: If $\sigma^n = u_i u_j \cdots u_k$ is an n -dimensional simplex of N , let $G_r(\sigma^n)$ be the set of points of N for which $x_i + x_j + \cdots + x_k > 1 - (r+1)(r+2)^{-1} \cdot 2^{-n-2}$, where x_i, x_j , etc., are barycentric coordinates. Let $\bar{G}_r(\sigma^n)$ be the set of points for which $x_i + x_j + \cdots + x_k \geq 1 - (r+1)(r+2)^{-1} \cdot 2^{-n-2}$. Let $G_\omega(\sigma^n)$ be the set of points for which $x_i + x_j + \cdots + x_k > 1 - 2^{-n-2}$, and let $\bar{G}_\omega(\sigma^n)$ be the set of points for which $x_i + x_j + \cdots + x_k \geq 1 - 2^{-n-2}$. The barycentric coordinates are continuous in N . Hence $G_r(\sigma^n)$ and $G_\omega(\sigma^n)$ are open while $\bar{G}_r(\sigma^n)$ and $\bar{G}_\omega(\sigma^n)$ are closed. If K is a closed subcomplex of N we define $G_r(K)$ to be the sum of the neighborhoods G_r of the simplexes of K , and similarly we define $\bar{G}_r(K)$, $G_\omega(K)$ and $\bar{G}_\omega(K)$.

The set $\bar{G}_r(K)$ is a closed set. For let p be an arbitrary point of $N - \bar{G}_r(K)$. It will be shown that p is contained in an open set U which does not intersect $\bar{G}_r(K)$. The point p is in some simplex of $N - K$, say $\sigma^n = u_k \cdots u_i \cdots u_j$. The closure $\bar{\sigma}^n$ of σ^n intersects K in a finite or empty complex $\bar{\sigma}^n \cdot K$. Let $U = G_r(\sigma^n) - \bar{G}_r(\bar{\sigma}^n \cdot K)$. The set $\bar{G}_r(\bar{\sigma}^n \cdot K)$, being a finite sum of closed sets, is closed and hence U is open. The point p is in σ^n and hence in $G_r(\sigma^n)$ but since p is not in $\bar{G}_r(K)$ it is not in $\bar{G}_r(\bar{\sigma}^n \cdot K)$. Hence p is contained in the open set U . Now suppose if possible that there is a point $p_1 \in U \cdot \bar{G}_r(K)$. Then $p_1 \in \bar{G}_r(\sigma^m)$ where σ^m is in K but not in $\bar{\sigma}^n \cdot K$. Let $\sigma^m = u_i \cdots u_j \cdots u_k$ and let $\bar{\sigma}^q$ be the intersection of $\bar{\sigma}^n$ and $\bar{\sigma}^m$, where $\sigma^q = u_i \cdots u_j$. Then $\sigma^q \subset \bar{\sigma}^n \cdot K$ and $q < n$ and $q < m$. Of

course σ^q may be empty. Thus p_1 is in $G_r(\sigma^n)$ and in $\tilde{G}_r(\sigma^m)$ but since $\sigma^q \subset \bar{\sigma}^n K$, p_1 is not in $\tilde{G}_r(\sigma^q)$. Hence the barycentric coordinates of p_1 satisfy the inequalities

$$x_h + \cdots + x_i + \cdots + x_j > 1 - \frac{r+1}{r+2} \cdot 2^{-n-2} \geq 1 - \frac{r+1}{r+2} \cdot 2^{-q-3},$$

$$x_i + \cdots + x_j + \cdots + x_k \geq 1 - \frac{r+1}{r+2} \cdot 2^{-m-2} \geq 1 - \frac{r+1}{r+2} \cdot 2^{-q-3},$$

$$x_i + \cdots + x_j < 1 - \frac{r+1}{r+2} \cdot 2^{-q-2}.$$

Adding the first two and subtracting the third gives

$$x_h + \cdots + x_i + \cdots + x_j + \cdots + x_k > 1,$$

which is impossible. Hence $U \cdot \tilde{G}_r(K) = 0$ and hence $\tilde{G}_r(K)$ is closed. Similarly $\tilde{G}_\omega(K)$ is closed. The sets $G_r(K)$ and $G_\omega(K)$ are sums of open sets and hence are open.

Let $V_i = St(u_i) - \tilde{G}_\omega(B_i)$. Thus V_i is an open set. We show that $\mathfrak{B} = \{V_i\}$ is a covering of N . For let p be any point of N . Then p is contained in some simplex $u_i \cdots u_k$, and therefore the barycentric coordinates of p satisfy the equality $x_i + \cdots + x_k = 1$. Let x_i be a largest coordinate of p , i. e., $x_i \geq x_j$ for each u_j of the simplex. Suppose, if possible, that p is not in V_i . Then p is in $\tilde{G}_\omega(u_h \cdots u_j)$ where $u_h \cdots u_j$ is some simplex not containing u_i . Let its dimension be n . Then $x_h + \cdots + x_j \geq 1 - 2^{-n-2}$. Since x_i is as large as any of the $n+1$ numbers x_h, \cdots, x_j , therefore

$$x_i \geq \frac{1}{n+1} \left(1 - \frac{1}{2^{n+2}}\right) = \frac{2^{n+2} - 1}{n+1} \cdot \frac{1}{2^{n+2}} > \frac{1}{2^{n+2}}.$$

Therefore $x_i + x_h + \cdots + x_j > 1$ which is impossible. Therefore p is in V_i . Therefore \mathfrak{B} is a covering of N .

Also \mathfrak{B} is a locally finite covering of N . For let p be any point of N . Let σ^n be the simplex of N which contains p . Then $G_\omega(\sigma^n)$ is an open set containing p . But $G_\omega(\sigma^n) \subset \tilde{G}_\omega(\sigma^n)$ which does not intersect V_i unless u_i is a vertex of σ^n . Hence $G_\omega(\sigma^n)$ intersects only a finite number of sets of \mathfrak{B} . Hence \mathfrak{B} is locally finite.

Let $H_0 = G_0(N^0)$, $H_1 = G_1(N^1)$ and, if $n > 1$, $H_n = G_n(N^n) - \tilde{G}_{n-2}(N^{n-2})$. We show that $\mathfrak{H} = \{H_n\}$ is a locally finite covering of N of order ≤ 2 . Let p be any point of N . Then p is in some simplex σ^m of N and hence in some set $\tilde{G}_m(N^m)$. Let the smallest m for which

$p \in \tilde{G}_m(N^n)$ be $m = n - 1$. Then $p \in \tilde{G}_{n-1}(\sigma^q)$ for some q -simplex σ^q with $q \leq n - 1$. Hence $p \in G_n(\sigma^q) \subset G_n(N^n)$. But p is not in $G_{n-2}(N^{n-2})$. Therefore $p \in H_n$. Therefore \mathfrak{S} is a covering of N . Also p is contained in the open set $G_n(N^n)$ which does not intersect H_m if $m \geq n + 2$. Hence \mathfrak{S} is a locally finite covering. Also, since $p \in \tilde{G}_{n-1}(N^{n-1})$, p is not in H_m for $m > n$ and, since p is not in $\tilde{G}_{n-2}(N^{n-2})$, p is not in H_m for $m \leq n - 2$. Hence p is in at most 2 sets of \mathfrak{S} and therefore \mathfrak{S} is of order ≤ 2 .

Let $W_{in} = V_i \cdot H_n$ and let $\mathfrak{B} = \{W_{in}\}$. Since \mathfrak{B} and \mathfrak{S} are locally finite coverings of N , \mathfrak{B} is a locally finite covering of N . Since $W_{in} \subset V_i \subset St(u_i)$, \mathfrak{B} is a refinement of the covering of N by the stars of its vertices. We show that the star of each vertex of $N(\mathfrak{B})$ the nerve of \mathfrak{B} , is finite dimensional. Let w_{in} be the vertex of $N(\mathfrak{B})$ corresponding to W_{in} . Let p be an arbitrary point of $W_{in} = V_i \cdot H_n$. Then $p \in V_i \cdot G_n(N^n)$ and therefore $p \in V_i \cdot G_n(\sigma^q)$ for some q -simplex with $q \leq n$. For each vertex u_k which is not a vertex of σ^q , $V_k \cdot G_n(\sigma^q) \subset V_k \cdot \bar{G}_\omega(\sigma^q) = 0$. Hence p is not in V_k . Hence p is in at most $q + 1$ of the sets of \mathfrak{B} . Therefore, since p is in at most two of the sets of \mathfrak{S} , p is in at most $2q + 2 \leq 2n + 2$ sets of \mathfrak{B} . Hence the star of w_{in} in $N(\mathfrak{B})$ has dimension at most $2n + 1$. This completes the proof of the lemma.

LEMMA 3.3. *If \mathfrak{U} is any locally finite covering of a normal space R there is a locally finite refinement \mathfrak{U}_2 of \mathfrak{U} such that the star of each vertex of the nerve of \mathfrak{U}_2 is finite dimensional.*

Proof. Let $\mathfrak{U} = \{U_i\}$ and let N be the nerve of \mathfrak{U} . By Theorem 1.1 there is a mapping ϕ of R in N such that, for each $U_i \in \mathfrak{U}$, $\phi^{-1}St(u_i) \subset U_i$. Let $\mathfrak{B} = \{W_{in}\}$ be a covering of N defined as in Lemma 3.2. Let $U_{in} = \phi^{-1}(W_{in})$ and let $\mathfrak{U}_2 = \{U_{in}\}$. Since \mathfrak{B} is locally finite so is \mathfrak{U}_2 . Since $U_{in} = \phi^{-1}W_{in} \subset \phi^{-1}St(u_i) \subset U_i$, \mathfrak{U}_2 is a refinement of \mathfrak{U} . Let $p \in U_{in}$. Then $\phi(p) \in W_{in}$, and hence $\phi(p)$ is in at most $2n + 2$ sets of \mathfrak{B} . Hence p is in at most $2n + 2$ sets of \mathfrak{U}_2 . Hence, in the nerve of \mathfrak{U}_2 , the star of the vertex corresponding to U_{in} has dimension at most $2n + 1$. This completes the proof.

Let K be a complex, $\bar{\sigma}^n$ a closed simplex of K and S^{n-1} the boundary of $\bar{\sigma}^n$. Then a mapping ϕ of a space R into K is called *essential* in $\bar{\sigma}^n$ if the partial mapping $\phi|_{\phi^{-1}S^{n-1}}$ cannot be extended to a mapping of $\phi^{-1}\bar{\sigma}^n$ into S^{n-1} . If the partial mapping can be extended ϕ is called *inessential* in $\bar{\sigma}^n$.

A locally finite covering \mathfrak{U} of a normal space R is called a *normal* covering if there is a mapping ϕ of R onto the nerve N of \mathfrak{U} such that each open set

of \mathcal{U} is the inverse image of the star of the corresponding vertex of N and such that ϕ is essential in every closed simplex of N .

LEMMA 3.4. *Each locally finite covering \mathcal{U} of a normal space R has a normal refinement which is star-finite if \mathcal{U} is star-finite and is finite if \mathcal{U} is finite.*

Proof. Let \mathcal{U} be any locally finite covering of R and let \mathcal{U}_1 be a refinement of \mathcal{U} such that the star of each vertex of the nerve N_1 of \mathcal{U}_1 , is finite dimensional. If \mathcal{U} is star-finite or finite, $\mathcal{U}_1 = \mathcal{U}$. Let ϕ_1 be a canonical mapping of R into N_1 .

We now define, by recursion, a sequence of subcomplexes $\{N_i\}$ of N_1 and a sequence of mappings $\{\phi_i\}$, where ϕ_i is a mapping of R into N_i . Suppose that, starting from N_1 and ϕ_1 , we have obtained a subcomplex N_i and a mapping ϕ_i of R into N_i which maps a neighborhood of each point of R continuously into a finite subcomplex of N_i . We call a simplex σ^n of N_i superfluous with respect to ϕ_i if ϕ_i is inessential in $\bar{\sigma}^n$ and if also $\bar{\sigma}^n$ is not on the boundary of a higher dimensional simplex of N_i . Since no superfluous simplex is on a boundary, if we remove all the superfluous simplexes of N_i we get a closed complex N_{i+1} . For each superfluous simplex $\bar{\sigma}^n$ of N_i let the partial mapping $\phi_i|_{\phi_i^{-1}\bar{\sigma}^n}$ be replaced by an extension of $\phi_i|_{\phi_i^{-1}(\bar{\sigma}^n - \sigma^n)}$ to a mapping of $\phi_i^{-1}\bar{\sigma}^n$ into $\bar{\sigma}^n - \sigma^n$. Let the resulting mapping of R into N_{i+1} be called ϕ_{i+1} . Because ϕ_{i+1} maps each of the closed sets $\phi_i^{-1}\bar{\sigma}^n$ continuously into $\bar{\sigma}^n$ or a subcomplex, therefore, if K_i is a finite subcomplex of N_i , ϕ_{i+1} maps $\phi_i^{-1}K_i$ continuously into K_i or a subcomplex. Hence, for each point p of R , a neighborhood mapped by ϕ_i into a finite subcomplex of N_i is mapped continuously by ϕ_{i+1} into a finite subcomplex of N_{i+1} .

Let p be a point of R and W a neighborhood of p mapped by ϕ_1 into a finite subcomplex K of N_1 . Then ϕ_i maps W continuously into $K \cdot N_i$. Since K is finite, and since $K \cdot N_{i+1}$ differs from $K \cdot N_i$ at most by omitting a finite number of simplexes of K , it follows that all but a finite number of the complexes $\{K \cdot N_i\}$ are identical. Hence all but a finite number of the mappings $\{\phi_i|_W\}$ are identical. Hence the sequence $\{\phi_i\}$ has a limit ϕ_ω which is a continuous mapping of R into $N_\omega = \prod_i N_i$.

Let $\bar{\sigma}^n$ be a closed simplex of N_ω and suppose if possible that ϕ_ω is inessential in $\bar{\sigma}^n$. Then clearly each ϕ_i is inessential in $\bar{\sigma}^n$. Hence, by Lemma 3.1, ϕ_i is inessential in each $(n+1)$ -simplex of N_i having $\bar{\sigma}^n$ on its boundary, and hence by recursion ϕ_i is inessential in every simplex of

$St_i(\bar{\sigma}^n)$, the star of $\bar{\sigma}^n$ in N_i . The star of each vertex, and hence of each simplex, of N_1 is finite dimensional. Let the dimension of $St_1(\bar{\sigma}^n)$ be m . Hence all m -simplexes of $St_1(\bar{\sigma}^n)$ are superfluous. Hence $\dim St_2(\bar{\sigma}^n) = m - 1$ and $\dim St_{m-n+1}(\bar{\sigma}^n) = n$. Hence $\bar{\sigma}^n$ is not in N_{m-n+2} and hence not in N_ω . This contradicts the assumption that $\bar{\sigma}^n$ is a closed simplex of N_ω . Hence ϕ_ω is essential in every simplex of N_ω .

It follows in particular that ϕ_ω is a mapping of R onto N_ω . We define \mathfrak{U}_ω to be the covering of R by the inverse images of the stars of the vertices of N_ω . Then, if we let each vertex u_i of N_ω correspond to the open set $U_{\omega i} = \phi_\omega^{-1}St_\omega(u_i)$ of \mathfrak{U}_ω , N_ω becomes the nerve of \mathfrak{U}_ω .

Thus ϕ_ω is a continuous mapping of R onto the nerve of \mathfrak{U}_ω , each open set of \mathfrak{U}_ω is the inverse image of the star of the corresponding vertex of N_ω , and ϕ_ω is essential in every closed simplex of N_ω . Therefore \mathfrak{U}_ω is a normal covering of R .

In the clearing off process, the set of points mapped into the star of a vertex never increases. Hence $U_{\omega i} = \phi_\omega^{-1}St_\omega(u_i) \subset \phi_1^{-1}St_1(u_i)$. Since ϕ_1 is canonical, $\phi_1^{-1}St_1(u_i) \subset U_{1i}$ where U_{1i} is the open set of \mathfrak{U}_1 corresponding to the vertex u_i of N_1 . Hence $U_{\omega i} \subset U_{1i}$. Therefore \mathfrak{U}_ω is a refinement of \mathfrak{U}_1 and hence a refinement of \mathfrak{U} . If \mathfrak{U} is star-finite, $\mathfrak{U}_1 = \mathfrak{U}$ is star-finite and, since \mathfrak{U}_ω is obtained by shrinking the open sets of \mathfrak{U}_1 , \mathfrak{U}_ω is also star-finite. Similarly if \mathfrak{U} is finite so is \mathfrak{U}_ω . This completes the proof of the lemma.

THEOREM 3.5. *If R is a normal space $\dim_L R = \dim_S R = \dim_F R$.*

Proof. It is sufficient to prove that each of the three inequalities, $\dim_L R \leq n$, $\dim_S R \leq n$, $\dim_F R \leq n$, implies both the others. This will be shown by proving that each of these inequalities holds if and only if, for each closed set R' of R , each mapping f of R' into the n -sphere S^n can be extended to a mapping F of R into S^n .

Let $\dim_L R \leq n$ or $\dim_S R \leq n$ or $\dim_F R \leq n$. Let R' be any closed set of R and let f be a mapping of R' into S^n . We assume that S^n is the boundary of an $(n+1)$ -simplex σ^{n+1} . By Tietze's Lemma f can be extended to a mapping g of R into $\bar{\sigma}^{n+1}$. Let \mathfrak{U}_1 be the covering of R by the inverse images of the stars of the vertices of $\bar{\sigma}^{n+1}$. The nerve N_1 of \mathfrak{U}_1 is identified with a subcomplex of $\bar{\sigma}^{n+1}$ so that g is a canonical mapping of R into N_1 . Let \mathfrak{U}_2 be a locally finite or star-finite or finite refinement of the finite covering \mathfrak{U}_1 such that the order of \mathfrak{U}_2 is $\leq n+1$. Let ϕ be a canonical

mapping of R into the nerve N_2 of \mathcal{U}_2 and let π be a projection¹³ from N_2 to N_1 . Since π is continuous on each finite subcomplex of N_2 $\pi\phi$ is a continuous mapping of R into N_1 . But, since N_2 has dimension $\leq n$, no simplex of N_2 is mapped into σ^{n+1} and hence $\pi\phi$ maps R into S^n . Let f_1 be the partial mapping $\pi\phi|_{R'}$. If p is a point of R' , $\phi(p)$ is in the closure of the simplex of N_2 determined by p , and π maps this closed simplex into the closure of the simplex containing $f(p)$. Hence the line segment $[f(p), f_1(p)]$ is contained in one of the closed simplexes of S^n . Let h be the mapping of $R' \times I$ into S^n which maps (p, t) into the point which divides the segment $[f(p), f_1(p)]$ in the ratio $t : 1 - t$. Then h is a uniform homotopy of f and f_1 . Therefore, by Theorem 2.1, since f_1 can be extended to a mapping $\pi\phi$ of R into S^n , f can be extended to a mapping F of R into S^n .

Now, conversely, let R be a normal space such that, for each closed set R' of R , each mapping f of R' into S^n can be extended to a mapping F of R into S^n . It follows that no mapping of R into a complex can be essential in any closed $(n+1)$ -simplex of the complex. Let \mathcal{U}_1 be any locally finite [star-finite, finite] covering of R . Let \mathcal{U}_2 be a locally finite [star-finite, finite] normal refinement of \mathcal{U}_1 (See Lemma 3.4). Then there is a mapping ϕ of R onto the nerve N_2 of \mathcal{U}_2 which is essential in every simplex of N_2 . Hence N_2 contains no $(n+1)$ -simplex. Hence the order of \mathcal{U}_2 is $\leq n+1$. Hence $\dim_L R \leq n$ and $\dim_S R \leq n$ and $\dim_F R \leq n$. This completes the proof.

We now define the *dimension* of a normal space R , $\dim R$, to be the common dimension $\dim_L R = \dim_S R = \dim_F R$.

In the proof of Theorem 3.5 we have proved the following:

COROLLARY¹⁴ 3.6. *If R is a normal space, $\dim R \leq n$ if and only if, for each closed set R' of R , each mapping f of R' into the n -sphere S^n can be extended to a mapping F of R into S^n .*

4. Čech cohomology groups. This section contains a summary of that

¹³ If a covering \mathfrak{B} is a refinement of a covering \mathfrak{U} , let each open set $V \in \mathfrak{B}$ be associated with some open set $U \in \mathfrak{U}$ containing V . This induces a mapping of the vertices of the nerve $N(\mathfrak{B})$ into the vertices of the nerve $N(\mathfrak{U})$. This vertex mapping can be extended to a barycentric mapping of the nerve of \mathfrak{B} in the nerve of \mathfrak{U} . Such a mapping of one nerve in another is called a projection ([8], p. 157). Barycentric mappings are defined and proved continuous on each finite or locally finite complex by S. Lefschetz ([24], p. 290). It can be shown that any barycentric mapping of one simplicial complex in another is continuous.

¹⁴ This theorem was proved by P. Alexandroff ([2], p. 170) for compact spaces. The extension to normal spaces was discovered independently by E. Hemmingsen.

part of the Čech cohomology theory which is needed in the rest of the paper. The family of coverings on which the cohomology groups are based is not necessarily a family of finite coverings but may be the family of all locally finite coverings or even the family of all coverings of the space. However, in the theorems of this paper, cohomology groups based on the family of all coverings are used only if this family is cofinal with the family of locally finite coverings, i. e., if the space is paracompact.

Let K be an arbitrary closure finite complex in which the boundary of the boundary of each cell is zero. Every simplicial complex satisfies these conditions. An n -chain C^n of K is an expression $a^\alpha \epsilon_\alpha^n$ where each a^α is an integer and each ϵ_α^n is an n -cell of K . Since K is closure finite the coboundary $\delta C^n = a^\alpha \eta_\alpha^n \epsilon_\alpha^{n+1}$ exists. An n -chain whose coboundary is zero is called an n -cocycle. The n -cocycles of K form an abelian group and the n -coboundaries form a subgroup of the n -cocycles. The difference group of the n -cocycles mod the n -coboundaries is called the n -cohomology group of K and is designated $H^n(K)$. In the Čech cohomology theory, which we shall now summarize, all complexes are assumed to be simplicial.

Let $K = \{\sigma_\alpha^n\}$ and $L = \{\tau_\beta^n\}$ be two arbitrary simplicial complexes and let f be a simplicial mapping of K into L . Let $f_\beta^{n\alpha} = 1, -1$, or 0 according as σ_α^n is mapped by f onto τ_β^n or $-\tau_\beta^n$ or is not mapped onto τ_β^n . If $C^n = a^\beta \tau_\beta^n$ is a chain of L let $f^* C^n = a^\beta f_\beta^{n+1, \alpha} \sigma_\alpha^{n+1}$. One easily verifies that $\eta_\gamma^\alpha f_\beta^{n\gamma} = \bar{\eta}_\beta^\gamma f_\gamma^{n+1, \alpha}$ where η_γ^α and $\bar{\eta}_\beta^\gamma$ are incidence numbers in K and L respectively. Therefore $\delta f^* C^n = a^\beta f_\beta^{n+1, \gamma} \eta_\gamma^\alpha \sigma_\alpha^{n+1} = a^\beta \bar{\eta}_\beta^\gamma f_\gamma^{n+1, \alpha} \sigma_\alpha^{n+1} = f^* \delta C^n$. Hence f^* induces a homomorphism of $H^n(L)$ into $H^n(K)$.

Let the vertices of K be ordered in an arbitrary fashion $u_0, u_1, \dots, u_n, \dots, u_\lambda, \dots$, and let $\{u'_\lambda\}$ and $\{u''_\lambda\}$ be two similar sequences. For each simplex σ_α^n of K we define $P\sigma_\alpha^n$ as follows: If $\sigma_\alpha^n = u_{\lambda_0} \dots u_{\lambda_n}$, where $\lambda_0 < \dots < \lambda_n$, then $P\sigma_\alpha^n = \Sigma_k (-1)^{k-1} u'_{\lambda_0} \dots u'_{\lambda_k} u''_{\lambda_k} \dots u''_{\lambda_n}$. The set of all simplexes appearing in the collection $\{P\sigma_\alpha^n\}$ together with all their faces, forms a simplicial complex which we denote by PK . The subcomplexes K' and K'' consisting of those simplexes of PK whose vertices have only single or double primes respectively are clearly isomorphic to K under the vertex correspondences $u_\lambda \leftrightarrow u'_\lambda$ and $u_\lambda \leftrightarrow u''_\lambda$. Thus to each chain C^n of K there correspond chains C'^n of K' and C''^n of K'' , and conversely.

LEMMA 4.1. *Let Γ^n be a cocycle of PK and let D_0^n and D_1^n be the parts of Γ^n on K' and K'' . Then the corresponding chains C_0^n and C_1^n of K are cohomologous in K .*

Proof. If $C^n = a^a \sigma_a^n$ is a finite chain and $D^n = b^a \sigma_a^n$ is any chain we denote by $KI(C^n; D^n)$ the integer $\Sigma_a a^a b^a$. We have $KI(\partial C^{n+1}; D^n) = \Sigma_a a^a \eta_\beta^a b^\beta = KI(C^{n+1}; \delta D^n)$. Hence since Γ^n is a cocycle $KI(\partial C^{n+1}; \Gamma^n) = 0$. One easily computes that $\partial P \sigma_a^n = \sigma''_a{}^n - \sigma'_a{}^n - P \partial \sigma_a^n$. Then $0 = KI(\partial P \sigma_a^n; \Gamma^n) = KI(\sigma''_a{}^n; \Gamma^n) - KI(\sigma'_a{}^n; \Gamma^n) - \Sigma_\beta \eta_\beta^a KI(P \sigma_\beta^{n-1}; \Gamma^n)$. Let $D_1^n = c_1^a \sigma''_a{}^n$ and $D_0^n = c_0^a \sigma'_a{}^n$. Then $KI(\sigma''_a{}^n; \Gamma^n) = c_1^a$ and $KI(\sigma'_a{}^n; \Gamma^n) = c_0^a$. Let $KI(P \sigma_\beta^{n-1}; \Gamma^n) = c^\beta$. Hence, for each α , $0 = c_1^\alpha - c_0^\alpha - \eta_\beta^\alpha c^\beta$. Let C^{n-1} be the chain $c^\beta \sigma_\beta^{n-1}$ in K . Then $\delta C^{n-1} = c^\beta \eta_\beta^a \sigma_a^n = c_1^a \sigma_a^n - c_0^a \sigma_a^n = C_1 - C_0$. Hence ^{14a} $C_1 \sim C_0$ in K , which proves the lemma.

Let R be a space. Let \mathcal{U}_α and \mathcal{U}_β be coverings of R and let N_α and N_β be their nerves. If \mathcal{U}_β is a refinement of \mathcal{U}_α there are one or more projections ¹⁵ of N_β into N_α . Let π_α^β and $\pi'_\alpha{}^\beta$ be two such projections. It can be shown, as for example by Čech ([8], p. 159), that there is a simplicial mapping f of PN_β into N_α such that $f|N'_\beta$ and $f|N''_\beta$ correspond to π_α^β and $\pi'_\alpha{}^\beta$. Let Γ_α^n be a cocycle of N_α . Then $f^* \Gamma_\alpha^n$ is a cocycle of $N_\beta \times I$. It follows from Lemma 4.1 that $\pi_\alpha^{*\beta} \Gamma_\alpha^n \sim \pi'^{*\beta} \Gamma_\alpha^n$ in N_β .

A collection $\Gamma^n = \{\Gamma_\alpha^n\}$ of one or more cocycles, one on each of the nerves of some of the coverings of R , is called a cocycle of R if the following condition is satisfied: If Γ_α^n and $\Gamma_\beta^n \in \Gamma^n$ there is a common refinement \mathcal{U}_γ of \mathcal{U}_α and \mathcal{U}_β , such that $\pi_\alpha^{*\gamma} \Gamma_\alpha^n \sim \pi_\beta^{*\gamma} \Gamma_\beta^n$ in N_γ . As shown above, the choice of the projection $\pi_\alpha^{*\gamma}$ and $\pi_\beta^{*\gamma}$ is irrelevant.

Two cocycles Γ_1^n and Γ_2^n of R are called cohomologous, $\Gamma_1^n \sim \Gamma_2^n$, if for each $\Gamma_{1\alpha}^n \in \Gamma_1^n$ and $\Gamma_{2\beta}^n \in \Gamma_2^n$ there is a common refinement \mathcal{U}_γ of \mathcal{U}_α and \mathcal{U}_β such that $\pi_\alpha^{*\gamma} \Gamma_{1\alpha}^n \sim \pi_\beta^{*\gamma} \Gamma_{2\beta}^n$ in N_γ . The sum $\Gamma_1^n + \Gamma_2^n$ of the cocycles of R is defined to be any cocycle Γ^n of R such that if $\Gamma_{1\alpha}^n \in \Gamma_1^n$, $\Gamma_{2\beta}^n \in \Gamma_2^n$, and $\Gamma_\gamma^n \in \Gamma^n$, then there is a common refinement \mathcal{U}_δ of \mathcal{U}_α , \mathcal{U}_β , and \mathcal{U}_γ such that $\pi_\gamma^{*\delta} \Gamma_\gamma^n \sim \pi_\alpha^{*\delta} \Gamma_{1\alpha}^n + \pi_\beta^{*\delta} \Gamma_{2\beta}^n$ in N_δ . The sum $\Gamma_1^n + \Gamma_2^n$ exists and is unique up to cohomology.

A cocycle Γ^n of R is determined up to cohomology by any $\Gamma_\alpha^n \in \Gamma^n$. Also $\Gamma^n \sim 0$ if for some $\Gamma_\alpha^n \in \Gamma^n$ there is a refinement \mathcal{U}_β of \mathcal{U}_α such that $\pi_\alpha^{*\beta} \Gamma_\alpha^n \sim 0$ in N_β .

The classes of cohomologous n -cocycles of R form a group $H^n(R)$, the n -dimensional cohomology group of R .

Let $Z = \{\mathcal{U}_\alpha\}$ be a family of coverings of R subject to the condition that

^{14a} For the convenience of the printer we depart from the Whitney [28] notation by using a tilde instead of a reversed tilde for "is cohomologous to." Since only cohomology is used in this paper, no misunderstanding need arise.

¹⁵ See note 13.

any two coverings \mathcal{U}_α and $\mathcal{U}_\beta \in Z$ have a common refinement $\mathcal{U}_\gamma \in Z$. A cocycle Γ^n of R is said to be in Z if for some $\Gamma_\alpha^n \in \Gamma^n$ the corresponding covering $\mathcal{U}_\alpha \in Z$. A cocycle Γ^n in Z is called ~ 0 in Z if for each $\Gamma_\alpha^n \in \Gamma^n$ for which the corresponding covering $\mathcal{U}_\alpha \in Z$ there is a refinement $\mathcal{U}_\beta \in Z$ of \mathcal{U}_α such that $\pi_\alpha^* \beta \Gamma_\alpha^n \sim 0$ in N_β . Let $Z \subset Z'$. Then if Γ^n is a cocycle in Z it is also in Z' , and, if $\Gamma^n \sim 0$ in Z , $\Gamma^n \sim 0$ in Z' .

The classes of n -cycles cohomologous in Z form a group $H_Z^n(R)$, the n -dimensional cohomology group of the family Z of coverings of R . In particular there are the cohomology groups $H_F^n(R)$, $H_S^n(R)$, and $H_L^n(R)$ of the families of finite, star-finite, and locally finite coverings respectively. The cohomology group $H_F^n(R)$ of the finite coverings of R is the usual Čech n -dimensional cohomology group of R .

If Z is a cofinal subfamily of Z' , i. e., if every covering $\mathcal{U} \in Z'$ has a refinement $\mathcal{B} \in Z$, then $H_Z^n(R) = H_{Z'}^n(R)$. If Z is a cofinal family of coverings of R , i. e., if every covering \mathcal{U} of R has a refinement $\mathcal{B} \in Z$ then $H_Z^n(R) = H^n(R)$.

Let R and S be two topological spaces and let f be a continuous mapping of R into S . If \mathcal{B}_α is any covering of S some of the open sets of \mathcal{B}_α are mapped by f^{-1} into open sets (not necessarily all distinct) which form a covering \mathcal{U}_α of R . Let ϕ be the barycentric of $N(\mathcal{U}_\alpha)$ into $N(\mathcal{B}_\alpha)$ which sends the vertex of $N(\mathcal{U}_\alpha)$ corresponding to the open set $f^{-1}(V) \in \mathcal{U}_\alpha$ into the vertex of $N(\mathcal{B}_\alpha)$ corresponding to $V \in \mathcal{B}_\alpha$. If $\Gamma^n = \{\Gamma_\alpha^n\}$ is any cocycle of S let $f^* \Gamma^n = \{\phi^* \Gamma_\alpha^n\}$. We show that $f^* \Gamma^n$ is a cocycle of R . Let \mathcal{B}_γ be any refinement of \mathcal{B}_α . If π_α^γ is a projection from $N(\mathcal{B}_\gamma)$ to $N(\mathcal{B}_\alpha)$ then $\pi_\alpha^\gamma \phi = \phi^{-1} \pi_\alpha^\gamma \phi$ is a projection from $N(\mathcal{U}_\gamma)$ to $N(\mathcal{U}_\alpha)$ such that $\phi \pi_\alpha^\gamma$ is the same mapping of $N(\mathcal{U}_\gamma)$ into $N(\mathcal{B}_\alpha)$ as $\pi_\alpha^\gamma \phi$. Therefore $\pi_\alpha^* \gamma \phi^* \Gamma_\alpha^n = \phi^* \pi_\alpha^* \gamma \Gamma_\alpha^n$. The inverse cocycle mapping $\pi_\alpha^* \gamma$ is determined up to cohomology independently of the choice of the projection. Hence, for each π_α^γ , $\pi_\alpha^* \gamma \phi^* \Gamma_\alpha^n \sim \phi^* \pi_\alpha^* \gamma \Gamma_\alpha^n$. Now if $\phi^* \Gamma_\alpha^n$ and $\phi^* \Gamma_\beta^n \in f^* \Gamma^n$, then Γ_α^n and $\Gamma_\beta^n \in \Gamma^n$ and hence there is a common refinement \mathcal{B}_γ of \mathcal{B}_α and \mathcal{B}_β such that $\pi_\alpha^* \gamma \Gamma_\alpha^n \sim \pi_\beta^* \gamma \Gamma_\beta^n$. Therefore $\pi_\alpha^* \gamma \phi^* \Gamma_\alpha^n \sim \phi^* \pi_\alpha^* \gamma \Gamma_\alpha^n \sim \phi^* \pi_\beta^* \gamma \Gamma_\beta^n \sim \pi_\beta^* \gamma \phi^* \Gamma_\beta^n$, which shows that $f^* \Gamma^n$ is a cocycle. Similarly using the fact that ϕ^* and π^* commute up to cohomology one can show that if $\Gamma^n \sim 0$ then $f^* \Gamma^n \sim 0$ and if $\Gamma^n = \Gamma'^n \pm \Gamma''^n$ then $f^* \Gamma^n = f^* \Gamma'^n \pm f^* \Gamma''^n$. Thus f induces a homomorphism of $H^n(S)$ into $H^n(R)$.

Let R' be a closed subspace of the space R . If \mathcal{U}_α is a covering of R the intersections with R' of the open sets of \mathcal{U}_α form a covering \mathcal{U}'_α of R' and every covering \mathcal{U}'_α of R' can be so obtained from a covering \mathcal{U}_α of R . The nerve $N'_\alpha = N(\mathcal{U}'_\alpha)$ can be considered as a closed subcomplex of $N'_\alpha = N(\mathcal{U}'_\alpha)$.

Let $\Gamma^n = \{\Gamma_a^n\}$ be a cocycle of R . If each Γ_a^n is in $N_a - N'_a$ and if the cohomologies $\pi^*_a \gamma \Gamma_a^n \sim \pi^*_\beta \gamma \Gamma_\beta^n$ can be satisfied in $N_\gamma - N'_\gamma$ we say that Γ^n is in $R - R'$. For such a Γ^n we say that $\Gamma^n \sim 0$ in $R - R'$ if the cohomologies $\pi^*_a \beta \Gamma_a^n \sim 0$ can be satisfied in $N_\beta - N'_\beta$.

Let $\Gamma'^n = \{\Gamma'_a{}^n\}$ be a cocycle of the space R' . Then $\Gamma'_a{}^n$ is a chain of N_a such that $\delta \Gamma'_a{}^n$ is in $N_a - N'_a$. The cocycle $\delta \Gamma'^n$ is determined up to cohomology in $N_a - N'_a$ by the cohomology class of $\Gamma'_a{}^n$. For if $\Gamma'_{1a}{}^n \sim \Gamma'_{2a}{}^n$ in N'_a there is a chain C^{n-1} in N'_a and a chain C^n in $N_a - N'_a$ such that $\delta C^{n-1} = \Gamma'_{1a}{}^n - \Gamma'_{2a}{}^n - C^n$ in N_a . Hence $0 = \delta \delta C^{n-1} = \delta \Gamma'_{1a}{}^n - \delta \Gamma'_{2a}{}^n - \delta C^n$. Hence $\delta \Gamma'_{1a}{}^n \sim \delta \Gamma'_{2a}{}^n$ in $N_a - N'_a$. We now show that $\Gamma^{n+1} = \{\delta \Gamma'_a{}^n\}$ is a cocycle of R in $R - R'$. Let $\pi'_a{}^\gamma$ be the projection from N'_γ to N'_a , etc. If $\delta \Gamma'_a{}^n$ and $\delta \Gamma'_\beta{}^n$ are in Γ^{n+1} then $\Gamma'_a{}^n$ and $\Gamma'_\beta{}^n$ are in Γ'^n . Hence there is a refinement \mathcal{U}'_γ of \mathcal{U}'_a and \mathcal{U}'_β such that $\pi'^*_a \gamma \Gamma'_a{}^n \sim \pi'^*_\beta \gamma \Gamma'_\beta{}^n$ in N'_γ . It can be assumed that \mathcal{U}'_γ is the intersection with R' of \mathcal{U}_γ , a refinement of \mathcal{U}_a and \mathcal{U}_β . Then $\delta \pi'^*_a \gamma \Gamma'_a{}^n \sim \delta \pi'^*_\beta \gamma \Gamma'_\beta{}^n$ in $N_\gamma - N'_\gamma$. Now $\pi^*_a \gamma \Gamma'_a{}^n = \pi'^*_a \gamma \Gamma'_a{}^n + C^n$ where C^n is a chain of $N_\gamma - N'_\gamma$ and hence $\delta \pi^*_a \gamma \Gamma'_a{}^n \sim \delta \pi'^*_a \gamma \Gamma'_a{}^n$ in $N_\gamma - N'_\gamma$. Therefore, since $\delta \pi^* = \pi^* \delta$, $\pi^*_a \gamma \delta \Gamma'_a{}^n \sim \pi^*_\beta \gamma \delta \Gamma'_\beta{}^n$ in $N_\gamma - N'_\gamma$. Therefore Γ^{n+1} is a cocycle of R in $R - R'$. We denote Γ^{n+1} by $\Psi(\Gamma'^n)$.

Let R' be a closed set of R and let S' be a closed set of S . If F is a continuous mapping of R into S which maps R' into S' then F^* can be seen to map cocycles of $S - S'$ into cocycles of $R - R'$ and to map cocycles which are cohomologous in $S - S'$ into cocycles which are cohomologous in $R - R'$. Thus F^* induces a homomorphism of the n -cohomology group of $S - S'$ into the n -cohomology group of $R - R'$. Let Δ'^n be any n -cocycle of S' and let \mathfrak{B}_a be a covering of S such that $\Delta'_a{}^n$ exists in the nerve of the corresponding covering \mathfrak{B}'_a of S' . The inverse images of the open sets of \mathfrak{B}_a and \mathfrak{B}'_a form coverings of R and R' with nerves N and N' such that $N'_a \subset N_a$. There are corresponding mappings Φ and $\phi = \Phi|_{N'_a}$ of N_a and N'_a into the nerves $N(\mathfrak{B}_a)$ and $N(\mathfrak{B}'_a) \subset N(\mathfrak{B}_a)$ respectively. Then $\Phi^* \Delta'_a{}^n = \phi^* \Delta'_a{}^n + C^n$, where C^n is a chain of $N - N'$. Hence $\delta \Phi^* \Delta'_a{}^n \sim \delta \phi^* \Delta'_a{}^n$ in $N - N'$. But $\delta \Phi^* \Delta'_a{}^n = \Phi^* \delta \Delta'_a{}^n$. Hence $\delta \phi^* \Delta'_a{}^n \sim \Phi^* \delta \Delta'_a{}^n$. Hence $\Psi^* \Delta'^n = \{\delta \phi^* \Delta'_a{}^n\} \sim \{\Phi^* \delta \Delta'_a{}^n\} = F^* \Psi \Delta'^n$ in $R - R'$. Thus, finally, $\Psi^* \Delta'^n \sim F^* \Psi \Delta'^n$ in $R - R'$.

5. Hopf's extension theorem. The n -dimensional cohomology group of the n -dimensional sphere S^n is cyclic infinite. Let Δ^n be a fixed cocycle in the generator of this group. We may assume that Δ^n is a cocycle of S^n in the family of finite coverings of S^n . We call Δ^n the fundamental cocycle of S^n .

LEMMA 5.1. *Let R be a normal space with $\dim R \leq n + 1$. Let R' be a*

closed subspace of R and let f be a mapping of R' into S^n . Let Z be a family of coverings of R such that all finite coverings of R are in Z and such that every covering in Z has a locally finite refinement. Then f can be extended to a mapping F of R into S^n if and only if $\Psi f^* \Delta^n \sim 0$ in $R - R'$ in Z .

Proof. 1) Let F be an extension of f . Let $\Delta^n = \{\Delta_a^n\}$. Then $F^* \Delta^n = \{\Phi^* \Delta_a^n\}$ is a cocycle of R and $f^* \Delta^n = \{\phi^* \Delta_a^n\}$ is a cocycle of R' . Let $\Delta_a^n \in \Delta^n$ and let \mathfrak{B}_a be a finite covering of S^n . Then $\mathfrak{U}_a = F^{-1} \mathfrak{B}_a$ is a finite covering of R . Then $\Phi^* \Delta_a^n$ is a cocycle of $N(\mathfrak{U}_a)$ and $\Phi^* \Delta_a^n = \phi^* \Delta_a^n + C_a^n$ where C_a^n is a chain in $N(\mathfrak{U}_a) - N(\mathfrak{U}_a)$. Therefore $\delta \phi^* \Delta_a^n = \delta(-C_a^n)$. This shows that $\Psi f^* \Delta^n \sim 0$ in $R - R'$ in the family of finite coverings of R . Therefore, since Z contains the finite coverings, $\Psi f^* \Delta^n \sim 0$ in $R - R'$ in Z .

2) Let f be a mapping of R' into S^n such that $\Psi f^* \Delta^n \sim 0$ in $R - R'$ in Z . We may assume that S^n is the boundary of an $(n+1)$ -simplex σ^{n+1} . Then S^n is a complex. Let \mathfrak{B}_0 be the covering of S^n by the stars of its vertices. We may assume that Δ^n is a cocycle of the generating element of $H^n(S^n)$ such that, for some cocycle Δ_0^n of $N(\mathfrak{B}_0)$, $\Delta_0^n \in \Delta^n$. Let $\mathfrak{U}'_0 = f^{-1}(\mathfrak{B}_0)$ and let \mathfrak{U}_0 be an extension of \mathfrak{U}'_0 to a covering of R . Then $N'_0 = N(\mathfrak{U}'_0)$ is isomorphic to a subcomplex of S^n . We identify N'_0 with this subcomplex so that $N'_0 \subset S^n$.

Let $\Gamma^n = f^* \Delta^n$. Let $\mathfrak{U}_\beta \in Z$ be a refinement of \mathfrak{U}_0 such that $\delta \pi^* \circ \beta \Gamma_0^n \sim 0$ in $N_\beta - N'_\beta$. Let \mathfrak{U}_a be a locally finite refinement of \mathfrak{U}_β of order $\leq n+2$. Then $\delta \pi_0^* \circ \alpha \Gamma_0^n \sim 0$ in $N_a - N'_a$. Therefore $\pi_0^* \circ \alpha \Gamma_0^n$ is part of a cocycle in N_a . The proof of Hopf's extension theorem for complexes ([28], p. 54) shows that the mapping $\pi_0^* \circ \alpha$ of N'_a into S^n can be extended to a mapping $\tilde{\pi}$ of N_a into S^n which is continuous on every finite subcomplex of N_a .

Let ϕ be a canonical mapping of R in N_a . Then $\tilde{\pi} \phi$ is a continuous mapping of R into S^n . Let f_0 be the partial mapping $\tilde{\pi} \phi|_{R'}$. Each point $x \in R'$ is mapped by ϕ in the closure of the simplex of N'_a determined by x . This closed simplex is mapped by $\tilde{\pi} = \pi_0^* \circ \alpha$ into the closure of the simplex of N'_0 determined by x , i. e., into the closure of the simplex of S^n containing $f(x)$. Therefore $f_0(x) = \tilde{\pi} \phi(x)$ is in the same closed simplex as $f(x)$. Let h be the mapping of $R' \times I$ in S^n which maps (x, t) in the point dividing the segment $[f_0(x), f(x)]$ in the ratio $t : 1 - t$. Then h is a uniform homotopy. Therefore by Borsuk's theorem (Theorem 2.1), since f_0 is uniformly homotopic to f , and since f_0 can be extended to a mapping $\tilde{\pi} \phi$ of R into S^n , f can be extended to a mapping F of R into S^n .

THEOREM 5.2. (*Hopf's extension theorem*) *Let R be a paracompact*

normal space with $\dim R \leq n + 1$. Let R' be a closed set of R and let f be a mapping of R' into the n -sphere S^n . Then f can be extended to a mapping F of R into S^n if and only if $\Psi f^* \Delta^n \sim 0$ in $R - R'$.

Proof. Since R is paracompact every covering of R has a locally finite refinement. If in Lemma 5.1 we let Z be the family of all open coverings of R we get Theorem 5.2 as an immediate consequence.

Hopf's extension theorem takes the following form if we use the ordinary Čech cocycles based on finite coverings.

THEOREM 5.3. (*Hopf's extension theorem with finite coverings*). Let R be a normal space with $\dim R \leq n + 1$. Let R' be a closed subspace of R and let f be a mapping of R' into the n -sphere S^n . Then f can be extended to a mapping F of R into S^n if and only if $\Psi f^* \Delta^n \sim 0$ in $R - R'$ in the family of finite coverings of R .

Proof. This theorem is obtained as an immediate consequence of Lemma 5.1 by letting Z be the family of finite coverings of R .

6. A class of coverings of $R \times I$. A certain class of coverings of the product space $R \times I$ plays an important role in the proof of Hopf's classification theorem. It is a generalization of the product covering which consists of the products of the sets of a covering of R by the sets of a covering of I . Let $\mathfrak{B} = \{V_{ai}\}$ be a covering of $R \times I$ by open sets each of which is a product set $V_{ai} = U_a \times W_{ai}$ of an open set of a locally finite covering $\mathfrak{U} = \{U_a\}$ by an open set (interval) of a covering \mathfrak{B}_a , corresponding to U_a , where $\mathfrak{B}_a = \{W_{ai}\}$ is an irreducible, and hence finite, covering of I by at least two intervals. It is assumed that for each irreducible covering, $\mathfrak{B}_a = \{W_{a0}, W_{a1}, W_{a2}, \dots, W_{ai}, \dots, W_{a,q(\alpha)}\}$, each two successive intervals intersect, no other pairs intersect, $0 \in W_{a0}$, and $1 \in W_{a,q(\alpha)}$ where $q(\alpha) \neq 0$. Such coverings \mathfrak{B} of $R \times I$ will be called *P-coverings*.

Let N be the nerve of a *P-covering* $\mathfrak{B} = \{V_{ai}\}$ and let v_{ai} be the vertex of N corresponding to $V_{ai} = U_a \times W_{ai}$. The intersection of \mathfrak{B} with $R \times 0$ and $R \times 1$ determine a covering $\mathfrak{U}' = \{U_a \times 0\}$ of $R \times 0$ and a covering $\mathfrak{U}'' = \{U_a \times 1\}$ of $R \times 1$ whose nerves N' and N'' we identify with the subcomplexes of N having vertices $\{v_{a0}\}$ and $\{v_{a,q(\alpha)}\}$ respectively. Under the correspondence of the vertices u_a of $N(\mathfrak{U})$ to v_{a0} and $v_{a,q(\alpha)}$, the complex $N(\mathfrak{U})$ corresponds to the isomorphic complexes N' and N'' , and the simplexes σ^n , chains C^n , etc. of $N(\mathfrak{U})$ correspond to the simplexes σ'^n , σ''^n , chains C'^n , C''^n , etc. of N' and N'' respectively.

Let the vertices of N be ordered as follows: Let $y_{\alpha i} \in I$ be chosen for each set $W_{\alpha i}$ so that, if $i = 0$, $y_{\alpha i} = 0$ and, if $i > 0$, $y_{\alpha i} \in W_{\alpha, i-1} \cdot W_{\alpha i}$. Thus for each α the points $y_{\alpha i}$ are ordered like the subscripts i . We say that $v_{\alpha i} < v_{\beta j}$ if either $y_{\alpha i} < y_{\beta j}$ or, when $y_{\alpha i} = y_{\beta j}$, $\alpha < \beta$.

Corresponding to each n -chain C^n of $N(\mathcal{U})$ we define a deformation chain DC^n of N as follows: Let $\sigma^n = u_\alpha u_\beta \cdots u_\gamma$. Then $D\sigma^n$ is defined to be the sum of all $n+1$ -simplexes σ^{n+1} of the form $v_{\delta, m-1} v_{\alpha i} v_{\beta j} \cdots v_{\delta m} \cdots v_{\gamma k}$ (where δ may be any of the subscripts $\alpha, \beta, \cdots \gamma$), such that for each vertex $v_{\epsilon r}$, with, $\epsilon \neq \delta$, $v_{\epsilon r} < v_{\delta m} < v_{\epsilon, r+1}$, the last relation being assumed trivially true if $v_{\epsilon, r+1}$ does not exist. Each such simplex exists in N if its vertices exist. For, since $v_{\epsilon r} < v_{\delta m}$, $y_{\epsilon r} \leq y_{\delta m}$ and, if $v_{\epsilon, r+1}$ exists, $y_{\delta m} \leq y_{\epsilon, r+1}$. Thus either $y_{\delta m}$ is between two points $y_{\epsilon r}$ and $y_{\epsilon, r+1}$ of the interval $W_{\epsilon r}$ or it is beyond a point $y_{\epsilon r}$ of a final interval $W_{\epsilon r} = W_{\epsilon, Q(\epsilon)}$. In either case $y_{\delta m} \in W_{\epsilon r}$. Let $p \in R$ be a point of the intersection $U_\alpha \cdot U_\beta \cdots U_\gamma$. Then $(p, y_{\delta m}) \in U_\epsilon \times W_{\epsilon r} = V_{\epsilon r}$. Also, since $y_{\delta m} \in W_{\delta, m-1} \cdot W_{\delta m}$, the point $(p, y_{\delta m}) \in V_{\delta, m-1} \cdot V_{\delta m}$. Hence $(p, y_{\delta m}) \in V_{\delta, m-1} \cdot V_{\alpha i} \cdot V_{\beta j} \cdots V_{\delta m} \cdots V_{\gamma k}$. Therefore the simplex $v_{\delta, m-1} v_{\alpha i} v_{\beta j} \cdots v_{\delta m} \cdots v_{\gamma k}$ exists in N . Thus $D\sigma^n$ is an $n+1$ -chain of N . If $C^n = a^\alpha \sigma_\alpha^n$ is an n -chain of $N(\mathcal{U})$, DC^n is defined to be the $n+1$ -chain $DC^n = a^\alpha D\sigma_\alpha^n$ of N .

LEMMA 6.1. *In the nerve of a P -covering of $R \times I$,*

$$\partial D\sigma^n = \sigma'^n - \sigma^n - D\partial\sigma^n.$$

Proof. Let $\sigma^n = u_\alpha u_\beta \cdots u_\gamma$ and let $\sigma^{n+1} = v_{\delta, m-1} v_{\alpha i} v_{\beta j} \cdots v_{\delta m} \cdots v_{\gamma k}$ be a simplex of $D\sigma^n$. Then $\partial\sigma^{n+1} = v_{\alpha i} v_{\beta j} \cdots v_{\delta m} \cdots v_{\gamma k} - v_{\alpha i} v_{\beta j} \cdots v_{\delta, m-1} \cdots v_{\gamma k} - v_{\delta, m-1} v_{\beta j} \cdots v_{\delta m} \cdots v_{\gamma k} + \cdots$. The first two simplexes of $\partial\sigma^{n+1}$, where all the subscripts α, β, \cdots of σ^n are represented, are called simplexes over σ^n , while the others are called simplexes over $\partial\sigma^n$. If $\sigma_1^n = v_{\alpha i} v_{\beta j} \cdots v_{\gamma k}$ is any n -simplex over σ^n such that, for each pair of its vertices $v_{\epsilon r}$ and $v_{\delta m}$, $v_{\delta m} < v_{\epsilon, r+1}$ and $v_{\epsilon r} < v_{\delta, m+1}$ then σ_1^n is on the boundary of exactly two simplexes of $D\sigma^n$ and with opposite signs, or $\sigma_1^n = \sigma'^n$ or $\sigma_1^n = \sigma''^n$. For, if $\sigma_1^n \neq \sigma'^n$, and if $v_{\delta m}$ is the largest vertex of σ_1^n (in the sense of $>$), then $m \neq 0$ and $v_{\delta, m-1} v_{\alpha i} \cdots v_{\delta m} \cdots v_{\gamma k}$ is the only simplex of $D\sigma^n$ having σ_1^n on its boundary with coefficient $+1$. And, if $\sigma_1^n \neq \sigma''^n$, there is a least, say $v_{\epsilon, r+1}$, of the vertices over vertices of σ^n such that $v_{\epsilon, r+1}$ is greater than all the vertices of σ_1^n , and $v_{\epsilon r} v_{\alpha i} \cdots v_{\epsilon, r+1} \cdots v_{\gamma k} = -v_{\epsilon, r+1} v_{\alpha i} \cdots v_{\epsilon r} \cdots v_{\gamma k}$ is the only simplex of $D\sigma^n$ having σ_1^n on its boundary with coefficient -1 . Hence if $\sigma_1^n \neq \sigma'^n$ and $\neq \sigma''^n$, σ_1^n has coefficient 0 in $\partial D\sigma^n$. And σ'^n is on the

boundary of only one simplex of $D\sigma^n$, where it has coefficient -1 , and σ''^n is on the boundary of only one simplex of $D\sigma^n$, where it has coefficient $+1$. Hence $\partial D\sigma^n = \sigma''^n - \sigma'^n + \text{simplexes over } \partial\sigma^n$.

If u_ϵ is the ρ -th vertex of σ^n then a typical simplex over $\partial\sigma^n$ of $\partial\sigma^{n+1}$ is $(-1)^\rho v_{\delta, m-1} v_{\alpha i} \cdots \tilde{v}_\epsilon \cdots v_{\delta m} \cdots v_{\gamma k}$ where v_ϵ means that the vertex with subscript ϵ is omitted. This is a simplex of the deformation chain of the oriented $n-1$ -simplex $(-1)^\rho u_\alpha \cdots \tilde{u}_\epsilon \cdots u_\delta \cdots u_\gamma$. The boundary of σ^n is $\partial\sigma^n = \tilde{u}_\alpha u_\beta \cdots u_\gamma - u_\alpha \tilde{u}_\beta \cdots u_\gamma + \cdots + (-1)^{\rho+1} u_\alpha \cdots \tilde{u}_\epsilon \cdots u_\gamma + \cdots$. Thus each simplex of $\partial\sigma^{n+1}$ over $\partial\sigma^n$ is a simplex of $D\partial\sigma^n$ with orientation reversed. Further if $(-1)^{\rho+1} v_{\delta, m-1} v_{\alpha i} \cdots \tilde{v}_\epsilon \cdots v_{\delta m} \cdots v_{\gamma k}$ is any simplex of $D\partial\sigma^n$ there is one and only one choice of r , namely the largest for which $a_{\epsilon r} < a_{\delta m}$, such that $v_{\delta, m-1} v_{\alpha i} \cdots v_{\epsilon r} \cdots v_{\delta m} \cdots v_{\gamma k}$ is a simplex of $D\sigma^n$. Therefore $\partial D\sigma^n = \sigma''^n - \sigma'^n - D\partial\sigma^n$, which completes the proof.

LEMMA 6.2. *If \mathfrak{B} is a P -covering over \mathfrak{U} of $R \times I$, and if D_0^n and D_1^n are the parts on N' and N'' of an n -cocycle Γ^n of $N(\mathfrak{B})$ then the corresponding cocycles C_0^n and C_1^n are cohomologous in $N(\mathfrak{U})$.*

Proof. Let $C_1^n = a''^a \sigma_a^n$ and $C_0^n = a'^a \sigma_a^n$. If σ_β^{n-1} is any $n-1$ -simplex of $N(\mathfrak{U})$, let $b^\beta = KI(D\sigma_\beta^{n-1}; \Gamma^n)$. Now since $\delta\Gamma^n = 0$, $0 = KI(D\sigma_a^n; \delta\Gamma^n) = KI(\partial D\sigma_a^n; \Gamma^n)$. Hence, by Lemma 6.1, $0 = KI(\sigma''^a; \Gamma^n) - KI(\sigma'^a; \Gamma^n) - KI(D\partial\sigma_a^n; \Gamma^n) = a''^a - a'^a - \eta_\beta^a b^\beta$. Therefore $a''^a \sigma_a^n - a'^a \sigma_a^n - \eta_\beta^a b^\beta \sigma_a^n = 0$ in $N(\mathfrak{U})$, i. e., $C_1^n - C_0^n - \delta b^\beta \sigma_\beta^{n-1} = 0$. Hence $C_1^n \sim C_0^n$ in $N(\mathfrak{U})$.

LEMMA 6.3. *Let R be a paracompact normal space and let Γ^n be an n -cocycle of $R \times I$. Then if D_0^n and D_1^n are the parts of Γ^n in $R \times 0$ and $R \times 1$ the corresponding cocycles C_0^n and C_1^n of R are cohomologous in R .*

Proof. Let $\Gamma^n = \{\Gamma_\alpha^n\}$ and let \mathfrak{B}_1 be a covering of $R \times I$ in which Γ_1^n is defined. We define a P -covering \mathfrak{B}_2 which is a refinement of \mathfrak{B}_1 . Each point $(x, t) \in R \times I$ is contained in a set $U \times W$ such that U is an open set R , W is an interval (open in I) of I , and $U \times W$ is contained in some open set of \mathfrak{B}_1 . The compact set $x \times I$ is covered by a finite number of these product sets $U_i(x) \times W_i(x)$. Let $\Pi_i U_i(x) = U(x)$. Then the sets $U(x) \times W_i(x)$ cover $x \times I$. It may be assumed that none of the intervals $W_i(x)$ is superfluous, i. e. that $\Sigma_{j \neq i} U_j(x) \neq I$. It may also be assumed that the sets $W_i(x)$ are divided if necessary so that no $W_i(x) \supset I$. The sets $U(x)$ form a covering of R . Since R is paracompact this covering has a locally finite refinement $\mathfrak{U} = \{U_\alpha\}$. Corresponding to each U_α choose x so

that $U_a \subset U(x)$ and let $W_{ai} = W_i(x)$. Then the sets $V_{ai} = U_a \times W_{ai}$ form a P -covering \mathfrak{B}_2 of $R \times I$ and \mathfrak{B}_2 is a refinement of \mathfrak{B}_1 .

We may assume that $\Gamma_2^n \in \Gamma^n$ is defined in the nerve N_2 of \mathfrak{B}_2 . Let $\mathfrak{U}_2 = \{U_a\}$ and let \mathfrak{U}'_2 and \mathfrak{U}''_2 be the corresponding coverings $\{U_a \times 0\}$ and $\{U_a \times 1\}$ of $R \times 0$ and $R \times 1$. Then $N'_2 = N(\mathfrak{U}'_2)$ and $N''_2 = N(\mathfrak{U}''_2)$ are subcomplexes of N_2 . We may assume that $D_{0,2}^n \in D_0^n$ and $D_{1,2}^n \in D_1^n$ are the part of Γ_2^n in N'_2 and N''_2 . Hence by Lemma 6.2 the corresponding cocycles of $N(\mathfrak{U}_2)$ are cohomologous. But the cocycles corresponding to $D_{0,2}^n$ and $D_{1,2}^n$ are $C_{0,2}^n \in C_0^n$ and $C_{1,2}^n \in C_1^n$. Hence $C_{0,2}^n \sim C_{1,2}^n$ in $N(\mathfrak{U}_2)$. Hence $C_0^n \sim C_1^n$ in R .

7. Hopf's classification theorem. Let $K \times I$ be the topological product of the simplicial complex K by the line interval I . Then $K \times I$ can be considered as a complex having as cells the simplexes $\sigma_a^n \times 0$ and $\sigma_a^n \times 1$ and the cells $\sigma_a^n \times J$ where σ_a^n is any simplex of K and J is the interior of the interval I . The closure of each cell of $K \times I$ is a finite complex and the boundary of the boundary of each cell is zero. Hence one can define chains, coboundaries and cocycles in $K \times I$. The complexes $K \times 0$ and $K \times 1$ are isomorphic to K under the correspondences $\sigma_a^n \leftrightarrow \sigma_a^n \times 0$ and $\sigma_a^n \leftrightarrow \sigma_a^n \times 1$. Thus to each chain C^n of K there correspond chains C'^n of $K \times 0$ and C''^n of $K \times 1$, and conversely.

LEMMA 7.1. *Let C_0^n and C_1^n be chains of an n -dimensional simplicial complex K and let D_0^n and D_1^n be the corresponding chains of $K \times 0$ and $K \times 1$. Then D_0^n and D_1^n are the parts in $K \times 0$ and $K \times 1$ of a cocycle Γ^n of $K \times I$ if and only if $C_0^n \sim C_1^n$ in K .*

Proof. Whitney's proof ([28], p. 53) of this lemma for the case of a finite complex carries over without change.

We now assume that S^n , ($n \geq 1$), is a simplicial complex, the boundary of an $n+1$ -simplex. We assume that the fundamental cocycle of S^n is $+\sigma_0^n$ where σ_0^n is a designated simplex of S^n .

We write ψ_0 and ψ_1 for the trivial mappings of K on $K \times 0$ and $K \times 1$: for $p \in K$, $\psi_0(p) = (p, 0)$, $\psi_1(p) = (p, 1)$.

LEMMA 7.2. *Let g_0 and g_1 be two barycentric mappings of an n -dimensional simplicial complex K into S^n . If $g_0^* \Delta^n \sim g_1^* \Delta^n$ in K then there is a mapping h of $K \times I$ in S^n continuous on each finite subcomplex of $K \times I$ and such that $h|K \times 0 = g_0 \psi_0^{-1}$, $h|K \times 1 = g_1 \psi_1^{-1}$.*

Proof. Projecting from a point inside σ_0^n one deforms g_0 and g_1 into normal mappings and then Whitney's proof ([28], p. 55) applies.

LEMMA 7.3. *Let R be a normal space with $\dim R \leq n$. Let Z be the family of locally finite coverings or the star-finite coverings or the finite coverings of R . If f_0 and f_1 are two mappings of R into the n -sphere S^n such that $f_0^* \Delta^n \sim f_1^* \Delta^n$ in Z then f_0 and f_1 are homotopic.*

Proof. Let $f_0^* \Delta^n = \Gamma^n = \{\Gamma_\alpha^n\}$ and $f_1^* \Delta^n = \Gamma''^n = \{\Gamma''_\alpha^n\}$. Let \mathfrak{B} be the covering of S^n by the stars of its vertices and let $\mathfrak{U}_0 = f_0^{-1} \mathfrak{B}$ and $\mathfrak{U}_1 = f_1^{-1} \mathfrak{B}$. Let \mathfrak{U}_2 be a common refinement in Z of \mathfrak{U}_0 and \mathfrak{U}_1 . Let \mathfrak{U}_3 be a refinement in Z of \mathfrak{U}_2 such that $\Gamma_3^n \sim \Gamma''_3^n$ in N_3 . Let \mathfrak{U}_4 be a refinement in Z of \mathfrak{U}_3 such that $\dim N_4 \leq n$. Let N_0 and N_1 be identified with the corresponding subcomplexes of S^n and let π_0^4 and π_1^4 be projections from N_4 to N_0 and N_1 respectively. Then π_0^4 and π_1^4 are simplicial mappings of N_4 into S^n and $\pi_0^{*4} \Delta^n \sim \Gamma_4^n \sim \Gamma''_4^n \sim \pi_1^{*4} \Delta^n$ in N_4 . Hence by Lemma 7.2 the corresponding mappings $\pi_0^4 \psi_0^{-1}$ of $N_4 \times 0$ and $\pi_1^4 \psi_1^{-1}$ of $N_4 \times 1$ can be extended to a mapping h of $N_4 \times I$ into S^n , with h continuous on each finite subcomplex of $N_4 \times I$. Let ϕ be a canonical mapping of R in N_4 . Then ϕ maps a neighborhood of each point $p \in R$ into a finite subcomplex K of N_4 and h maps $K \times I$ continuously in S^n . Hence $h\phi$ is continuous and hence the mappings $\pi_0^4 \phi$ and $\pi_1^4 \phi$ of R in S^n are homotopic. Any point $p \in R$ is mapped by ϕ into the closure of the simplex $\sigma_4(p)$ determined by p in N_4 and this is mapped by π_0^4 into the closure of the simplex determined by p in N_0 , a closed simplex which also contains $f_0(p)$. Hence $\pi_0^4 \phi$ is uniformly homotopic to f_0 . Similarly $\pi_1^4 \phi$ is uniformly homotopic to f_1 . Therefore f_0 is homotopic to f_1 , which completes the proof.

LEMMA 7.4. *Let R be a normal space with $\dim R \leq n$. Let Z be the family of locally finite coverings or the star-finite coverings or the finite coverings of R . Then, from the correspondence between the cocycle $f^* \Delta^n$ and the mapping f , each element of the n -cohomology group of R in Z corresponds to one and only one class of homotopic mappings of R into the n -sphere S^n .*

Proof. By Lemma 7.3 if two mappings correspond to cocycles of the same cohomology element they are homotopic and hence each cohomology element corresponds to at most one homotopy class. Hence it is sufficient to show that to each cohomology element corresponds some mapping. Thus we must show that if Γ^n is an n -cocycle of Z there is a mapping f of R in S^n such that $f^* \Delta^n \sim \Gamma^n$ in Z .

Let $\Gamma^n = \{\Gamma_a^n\}$ be any n -cocycle of Z and let \mathcal{U}_1 be a normal covering of R in Z such that Γ_1^n is defined in the at most n -dimensional nerve N_1 of \mathcal{U}_1 . Let g_1 be a mapping of N_1 into S^n which maps the $n-1$ dimensional part N_1^{n-1} of N_1 into a designated vertex p_0 of S^n and which maps each n -simplex σ_a^n of N_1 so that, for some simplicial subdivision σ'_a^n of σ_a^n in which the boundary $\delta\sigma_a^n$ is not subdivided, σ'_a^n is mapped simplicially and exactly $|a^n|$ n -simplexes of σ'_a^n are mapped on a designated n -simplex σ_0^n of S^n , the mapping preserving orientation on these simplexes if a^n is positive, otherwise reversing orientation. The construction of such a subdivision and mapping is easy and omitted. Let N_2 be the resulting subdivision of N_1 and g_2 the resulting simplicial mapping of N_2 in S^n . Let ϕ_2 be the mapping of R on N_2 corresponding to a canonical mapping ϕ_1 of R on N_1 . Let f be the mapping $g_1\phi_1 = g_2\phi_2$ of R in S^n . Let $f^*\Delta^n = C^n$ in Z .

If \mathfrak{B}_1 is the covering of S^n by the stars of its vertices, the complex S^n is the nerve of \mathfrak{B}_1 . We assume that the cocycle of Δ^n on this nerve is $\Delta_1^n = +\sigma_0^n$. Let $\mathcal{U}_3 = f^{-1}(V_1)$. Its nerve N_3 is S^n or a subcomplex of S^n . Let ψ be the identical mapping of N_3 in S^n . Let \mathcal{U}_2 be the inverse image by ϕ_2 of the covering of N_2 by the stars of its vertices. Then \mathcal{U}_2 is in Z , its nerve is N_2 and it is a common refinement of \mathcal{U}_1 and \mathcal{U}_3 . The mapping $\pi_3^2 = \psi^{-1}g_2$ is a projection from N_2 to N_3 . Let π_1^2 be a projection from N_2 to N_1 which maps exactly one simplex of σ'_a^n into σ_a^n . Then $\pi_1^{*2}\Gamma_1^n$ has coefficient a^n on one simplex of σ'_a^n and $\pi_3^{*2}C_3^n$ has coefficient $\text{sgn } a^n$ on $|a^n|$ simplexes of σ'_a^n . Hence $\pi_3^{*2}C_3^n \sim \pi_1^{*2}\Gamma_1^n$ in N_2 . Hence $C^n \sim \Gamma^n$ in Z . Therefore $f^*\Delta^n \sim \Gamma^n$ in Z , which completes the proof.

THEOREM 7.5. (Hopf classification theorem). *Let R be a paracompact normal space with $\dim R < n$, ($n \geq 1$). Then the correspondence between the cocycle $f^*\Delta^n$ and the mapping f induces a 1—1 correspondence between the elements of the n -cohomology group of R and the classes of homotopic mappings of R into the n -sphere S^n .*

Proof. The family Z of locally finite coverings of R forms a cofinal family and hence each element of $H^n(R)$ determines a unique element of $H_Z^n(R)$ which by Lemma 7.4 corresponds to a unique homotopy class of mappings of R in S^n .

Conversely, let f_0 and f_1 be elements of the same homotopy class of mappings of R in S^n . Let h be a mapping of $R \times I$ in S^n such that for $x \in R$, $h(x, 0) = f_0(x)$, $h(x, 1) = f_1(x)$. Then $f_0^*\Delta^n$ and $f_1^*\Delta^n$ are the cocycles

of R corresponding to the parts in $R \times 0$ and $R \times 1$ of the cocycle $h^* \Delta^n$ in $R \times I$. Hence, by Lemma 6.3, $f_0^* \Delta^n \sim f_1^* \Delta^n$ in R . Thus, to each homotopy class of mappings of R in S corresponds a unique element of $H^n(R)$. This completes the proof.

8. Bruschi's theorem. In the special case of mappings into a circumference S^1 , restrictions on the dimension of the space R can be removed. This is a consequence of the fact that every mapping of an n -sphere, $n > 1$, in a circumference S^1 is inessential. We briefly indicate the changes necessary to prove the more general mapping theorem when the image sphere is S^1 .

a) Let g_0 and g_1 be two mappings of a simplicial complex K into the circumference S^1 , each continuous on each finite subcomplex of K , such that all vertices of K are mapped by g_0 and g_1 into a designated point p_0 of S^1 . If the degrees of the mappings g_0 and g_1 for the simplex σ_a^1 of K are a^a and b^a respectively and if $C_0^1 = a^a \sigma_a^1$ and $C_1^1 = b^a \sigma_a^1$ are cohomologous cocycles of K , then there is a mapping h of $K \times I$ in S^1 , continuous on each finite subcomplex of $K \times I$, and such that, if $x \in K$, $h(x, 0) = g_0(x)$ and $h(x, 1) = g_1(x)$.

By Lemma 7.1 there is a cocycle $\Gamma^1 = a^a \sigma_a^1 + b^a \sigma_a^1 + c^\beta (\sigma_\beta^0 \times I)$ of $K \times I$ whose parts on $K \times 0$ and $K \times 1$ correspond to C_0^1 and C_1^1 . Let the partial mapping by h of $K \times 0$ and $K \times 1$ be defined by $h(x, 0) = g_0(x)$ and $h(x, 1) = g_1(x)$, and let the mapping be extended to the remaining 1-cells of $K \times I$ by mapping $\sigma_\beta^0 \times I$ with degree c^β . Since Γ^1 is a cocycle, the boundary of each 2-cell $\sigma_a^1 \times I$ is mapped with degree zero, and the mapping can be extended over the 2-cells. The boundary of each 3-cell is now mapped in S^1 , necessarily inessentially, and hence the mapping can be extended to all 3-cells. And so on. The mapping h is thus defined on all cells of $K \times I$.

b) Let g_0 and g_1 be two simplicial mappings of a simplicial complex K in the boundary S^1 of a 2-simplex. If $g_0^* \Delta^1 \sim g_1^* \Delta^1$ in K there is a mapping h of $K \times I$ in S^1 , continuous on each finite subcomplex of $K \times I$, and such that, if $x \in K$, $h(x, 0) = g_0(x)$ and $h(x, 1) = g_1(x)$.

We assume that $\Delta^1 = + \sigma_0^1$ where σ_0^1 is a designated simplex of S^1 . Let p_0 be a point of S^1 not in σ_0^1 . Projecting from a point of σ_0^1 one deforms the mappings g_0 and g_1 so that all vertices are mapped into p_0 . Thus, b) is seen to be a consequence of a).

c) Let R be a normal space and let Z be the family of locally finite coverings or the star-finite coverings or the finite coverings of R . If f_0 and f_1

are mappings of R in the circumference S^1 such that $f_0^* \Delta^1 \sim f_1^* \Delta^1$ in Z then f_0 and f_1 are homotopic.

This follows from b) using the method of proof of Lemma 7.3.

d) If Γ^1 is a 1-cocycle of R in Z there is a mapping f of R in S^1 such that $f^* \Delta^1 \sim \Gamma^1$ in Z .

Let \mathfrak{U}_1 be a covering of R in which $\Gamma_1^1 = a^a \sigma_a^1$ is defined. Let ϕ_1 be a canonical mapping of R in the nerve N_1 of \mathfrak{U}_1 . Let g_1 be the following mapping of N_1 in S^1 . Let each vertex of N_1 be mapped in a designated vertex p_0 . Let each 1-simplex σ_a^1 of N_1 be subdivided so it can be mapped barycentrically in S^1 with degree a^a . We can assume that exactly $|a^a|$ simplexes of the subdivision $\sigma'_a{}^1$ of σ_a^1 are mapped in a designated 1-simplex σ_0^1 of S^1 . Since Γ_1^1 is a cocycle the boundary of each 2-simplex of N_1 is mapped with degree zero. Hence the mapping can be extended over each 2-simplex. This extended mapping can be deformed into a simplicial mapping of a subdivision $\sigma'_\beta{}^2$ of σ_β^2 in such a way that the subdivisions $\sigma'_a{}^1$ of simplexes σ_a^1 of the boundary and the mappings of the boundary simplexes are unchanged. Since each boundary of a 3-simplex is mapped inessentially in S^1 the mapping can be extended to each 3-simplex and deformed into a suitable simplicial mapping, and so on. Thus we get a simplicial mapping g_2 of a subdivision N_2 of N_1 into S^1 . If ϕ_2 is the mapping of R in N_2 corresponding to the mapping ϕ_1 in N_1 , then $f = g_1 \phi_1 = g_2 \phi_2$ is a continuous mapping of R in S^1 . As in the proof of Lemma 7.4 one proves that $f^* \Delta^1 \sim \Gamma^1$ in Z .

e) If R is a paracompact normal space, the correspondence between the cocycle $f^* \Delta^1$ and the mapping f induces a 1 — 1 correspondence between the elements of the 1-cohomology group of R and the classes of homotopic mappings of R into the circumference S^1 .

This follows from d) and Lemma 6.3.

The mappings into a circumference form a group. Thus, if we identify S^1 with the additive group of real numbers mod 1, we define the sum $f_0 + f_1$ of the mappings f_0 and f_1 of R in S^1 by $(f_0 + f_1)(p) = f_0(p) + f_1(p) \bmod 1$ for each $p \in R$. The inessential mappings of R in S form a subgroup. The difference group of the mappings of R in S^1 mod the inessential mappings is a group whose elements are the classes of homotopic mappings of R in S^1 .

Let Γ^1, C^1, D^1 be 1-cocycles of a paracompact normal space R such that $\Gamma^1 \sim C^1 + D^1$. Let U_1 be a locally finite covering of R such that $\Gamma_1^1 \sim C_1^1 + D_1^1$ in N_1 . We may assume in fact that $\Gamma_1^1 = C_1^1 + D_1^1$ in N_1 . Let $\Gamma_1^1 = a^a \sigma_a^1, C_1^1 = c^a \sigma_a^1, D_1^1 = d^a \sigma_a^1$. Then $a^a = c^a + d^a$. Let g_0, g_1, g_2

be simplicial mappings of subdivisions of N_1 in S^1 mapping the vertices of N_1 into a vertex p_0 of S^1 and mapping subdivisions of σ_a^1 in S^1 with degrees a^a , c^a and d^a respectively. Then $g_1 + g_2$ maps σ_a^1 in S^1 with the same degree $a^a = c^a + d^a$ as g_0 . Hence, by a), there is a mapping h of $N_1 \times I$ in S^1 , continuous on each finite subcomplex, such that, if $x \in N_1$, $h(x, 0) = g_0(x)$ and $h(x, 1) = (g_1 + g_2)(x)$. Let ϕ be a canonical mapping of R in N_1 . Let $f_0 = g_0\phi$, $f_1 = g_1\phi$, $f_2 = g_2\phi$. Then $f_0^*\Delta^1 \sim \Gamma^1$, $f_1^*\Delta^1 \sim C^1$, and $f_2^*\Delta^1 \sim D^1$. And $f_1 + f_2 = (g_1 + g_2)\phi$ is homotopic to $g_0\phi = f_0$. Hence, under the correspondence between $f^*\Delta^1$ and f , the sum of the cocycles corresponds to the sum of the corresponding mappings. Thus we have the following form of Bruschlinsky's Theorem:

THEOREM 8.1. (*Bruschlinsky's Theorem*). *If R is a paracompact normal space the correspondence between the cocycle $f^*\Delta^1$ and the mapping f induces an isomorphism between the 1-cohomology group of R and the group of homotopy classes of mappings of R into the circumference S^1 .*

9. Finite coverings. A compact space is a paracompact space in which the family of finite coverings forms a cofinal family. Hence the Hopf theorem and Bruschlinsky's theorem hold for compact normal spaces even if cocycles and cohomologies are based on the family of finite coverings. Some similar theorems stated in terms of finite coverings hold, also, for non-compact spaces. We shall return to consideration of these theorems after a digression.

We first show that homotopic mappings of a countably compact space, or, in particular a compact space, into a metric space are uniformly homotopic.

LEMMA ¹⁰ 9.1. *Let R be a countably compact space, T a compact space in which the first countability axiom is satisfied, and \mathcal{U} a countable or finite covering of $R \times T$ by open sets. Then there exist finite coverings \mathfrak{B} and \mathfrak{B} of R and T respectively such that $\mathfrak{B} \times \mathfrak{B}$ is a refinement of \mathcal{U} .*

Proof. Since R is countably compact so is $R \times t$ and hence, because $R \times t$ is covered by the countable family \mathcal{U} of open sets, it is covered by a finite number of these, say $\{U_1(t), \dots, U_r(t)\}$. We define in succession open sets V_i^t of R and W_i^t of T , such that $V_i^t \times W_i^t \subset U_i(t)$ and such that

$$\{V_1^t \times W_1^t, \dots, V_{j-1}^t \times W_{j-1}^t, U_j(t), \dots, U_r(t)\}$$

¹⁰ A similar theorem was proved by Čech ([9], p. 3).

covers $R \times t$. Suppose that V_i^t, W_i^t have been defined for $i < j$. We must define V_j and W_j so that $F_j^t \subset V_j^t \times W_j^t \subset U_j^t$ where F_j^t is the closed set of $R \times t$: $F_j^t = R \times t - (R \times t) \cdot [V_1^t \times W_1^t + \dots + V_{j-1}^t \times W_{j-1}^t + U_{j+1}(t) + \dots + U_r(t)]$. For each point $(x, t) \in F_j^t$ there are open sets $V_j(x, t)$ of R and $W_j(x, t)$ of T such that $(x, t) \in V_j(x, t) \times W_j(x, t) \subset U_j(t)$. Let $\{N_i^n\}$ be a complete family of neighborhoods of t in T such that $N_i^n \supset N_{i+1}^n$. If $G_{jn}^t = \sum V_j(x, t)$, where the summation is over all x such that $W_j(x, t) \supset N_i^n$, then G_{jn}^t is an open set of R , $G_{jn}^t \times N_n^t \subset U_j(t)$, and $G_{jn}^t \subset G_{j,n+1}^t$. Since every $W_j(x, t)$ contains some neighborhood N_i^n , therefore $\sum_{n=1}^{\infty} G_{jn}^t \times t \supset F_j^t$. Therefore, since F_i^t is countably compact, there is an n such that $G_{jn}^t \times t \supset F_j^t$. Let this set G_{jn}^t be called V_j^t and let the corresponding neighborhood N_i^n be called W_j^t . Then, as required, $F_j^t \subset V_j^t \times W_j^t \subset U_j(t)$.

Thus eventually $R \times t$ is covered by the sets $\{V_i^t \times W_i^t\}$, where $\mathfrak{B}(t) = \{V_i^t\}$ is a finite covering of R . Let $W(t) = \Pi_i W_i^t$. A finite number of the open sets $W(t)$ cover the compact space T . Let this finite covering of T be $\mathfrak{B} = \{W(t_k)\}$. Let \mathfrak{B} be a common refinement of the finite coverings $\mathfrak{B}(t_k)$. Then $\mathfrak{B} \times \mathfrak{B}$ is a finite covering of $R \times T$. Each open set of the covering $\mathfrak{B} \times \mathfrak{B}$ is, for some t , of the form $V \times W(t)$ where V is contained in an open set V_i^t of $\mathfrak{B}(t)$ and $V_i^t \times W(t) \subset V_i^t \times W_i^t \subset U_i(t)$ which is an open set of \mathfrak{U} . Thus $\mathfrak{B} \times \mathfrak{B}$ is a refinement of \mathfrak{U} . This completes the proof.

It is an immediate consequence that, if R and T satisfy the hypotheses of Lemma 9.1, $R \times T$ is countably compact.

LEMMA 9.2. *Two continuous mappings f_0 and f_1 of a countably compact space R in a metric space S are uniformly homotopic if and only if they are homotopic.*

Proof. It follows immediately from the definitions that if f_0 and f_1 are uniformly homotopic they are homotopic.

Let f_0 and f_1 be homotopic. Then there is a continuous mapping h of $R \times I$ in S such that $h(x, 0) = f_0(x)$, $h(x, 1) = f_1(x)$. Since R is countably compact, and I is compact and satisfies the first countability axiom, $R \times I$ is countably compact. Since $h(R \times I)$ is the continuous image of a countably compact space it is a countably compact subspace of S and hence a compact metric space. Hence, for any $\epsilon > 0$, $h(R \times I)$ has a finite covering by open sets, each of diameter $< \epsilon$. Their inverse images form a finite covering of

$R \times I$ by open sets. Let this covering be \mathfrak{U} . Then, by Lemma 9.1, there are coverings \mathfrak{B} and \mathfrak{B} of R and I respectively such that $\mathfrak{B} \times \mathfrak{B}$ is a refinement of \mathfrak{U} , and hence such that $h(V \times W)$ has diameter $< \epsilon$ for any $V \in \mathfrak{B}$ and $W \in \mathfrak{B}$.

Since I is compact there is a number $\delta > 0$ such that for every pair of points t, t' of I such that $|t - t'| < \delta$, there is an open set $W \in \mathfrak{B}$ which contains both points. Hence, if $|t - t'| < \delta$, (x, t) and (x, t') are in the same set $V \times W$, and hence $\rho(h(x, t), h(x, t')) < \epsilon$. Therefore f_0 and f_1 are uniformly homotopic, which completes the proof.

We now recall some properties of the Tychonoff-Čech compactification of a normal space. Čech [10] has shown that if R is a normal space there is a compact Hausdorff space $\beta(R)$ and a mapping ψ of R into $\beta(R)$ such that:

- 1) The image set ψR is a normal Hausdorff space which is dense in $\beta(R)$.
- 2) Any bounded continuous real function f on R has the form $f(x) = F\psi(x)$ where F is a similar function on $\beta(R)$. By 1) the function F is unique.
- 3) It follows from 2) that any mapping f of R into S^n determines a mapping F of $\beta(R)$ into S^n , such that $f(x) = F\psi(x)$.

We need these facts in the proof of

THEOREM 9.3. *Let R be a normal space with $\dim R \leq n$, ($n \geq 1$). The correspondence between the cocycle $f^* \Delta^n$ and the mapping f induces a 1—1 correspondence between the elements of the n -cohomology group $H_F^n(R)$ based on finite coverings and the classes of uniformly homotopic mappings of R into the n -sphere S^n .*

Proof. Corresponding to each mapping f of R in S^n is a mapping F of $\beta(R)$ in S^n such that $f = F\psi$. This mapping f determines a cocycle $f^* \Delta^n = \psi^* F^* \Delta^n$. Thus, corresponding to a mapping f is a mapping F and the cocycle $F^* \Delta^n$ as well as the cocycle $f^* \Delta^n$. Since $\beta(R)$ is compact, and hence paracompact, the correspondence between F and $F^* \Delta^n$ leads to a 1—1 correspondence between the homotopy classes of mappings of $\beta(R)$ in S^n and the elements of the n -cohomology group of $\beta(R)$. It is sufficient to show that a) the correspondence between f and F leads to a 1—1 correspondence between the uniform homotopy classes of mappings of R in S^n and the homotopy classes of mappings of $\beta(R)$ in S^n , and that b) the mapping ψ^* of the cocycles of $\beta(R)$ into the cocycles of R induces an isomorphism between the cohomology group $H^n(\beta(R))$ and the cohomology group $H_F^n(R)$.

a) First let F_0 and F_1 be homotopic mappings of $\beta(R)$ in S^n , and let f_i map R in S^n so that $f_i(x) = F_i\psi(x)$. Since $\beta(R)$ is compact, F_0 and F_1 are uniformly homotopic (by Lemma 9.2) and hence the same is true of f_0 and f_1 .

On the other hand let f_0 and f_1 be uniformly homotopic mappings of R in S^n and let F_0, F_1 be the corresponding mappings of $\beta(R)$. Let $g(x, t)$ be the mapping of $R \times I$ in S^n which gives the uniform homotopy of f_0 and f_1 . For fixed t , $g(x, t)$ is a continuous mapping of R in S^n which we call $g_t(x)$ and there is a corresponding continuous mapping G_t of $\beta(R)$ in S^n such that $G_t\psi(x) = g_t(x)$. Defining $G(y, t)$ to be $G_t(y)$ for $y \in \beta(R)$ we have a transformation of $\beta(R) \times I$ into S^n with $G(y, 0) = F_0(y)$, $G(y, 1) = F_1(y)$. Hence, if we can show that $G(y, t)$ is continuous, F_0 and F_1 are homotopic.

Let $\epsilon > 0$, and let $(y_0, t_0) \in \beta(R) \times I$. Since G_{t_0} is continuous there is a neighborhood U of y_0 in $\beta(R)$ such that, for $y' \in U$, the distance $\rho(G_{t_0}(y_0), G_{t_0}(y')) < \epsilon/3$. Since $g(x, t)$ is a uniform homotopy there is a δ -neighborhood W of t_0 such that, if $x \in R$ and $t \in W$, $\rho(g_{t_0}(x), g_t(x)) = \rho(G_{t_0}\psi(x), G_t\psi(x)) < \epsilon/3$. Now let (y, t) be any point of the neighborhood $U \times W$ of (y_0, t_0) in $\beta(R) \times I$. Since G_t is continuous there is a neighborhood $U_1 \subset U$ of y such that, if $y' \in U_1$, $\rho(G_t(y'), G_t(y)) < \epsilon/3$.

But since ψR is dense in $\beta(R)$ there is a point $x \in R$ such that $\psi(x) \in U_1$. Hence $\rho(G(y_0, t_0), G(y, t)) \leq \rho(G_{t_0}(y_0), G_{t_0}\psi(x)) + \rho(G_{t_0}\psi(x), G_t\psi(x)) + \rho(G_t\psi(x), G_t(y)) < \epsilon$. Hence F_0 and F_1 are homotopic. Hence we have a 1—1 correspondence between the uniform homotopy classes of mappings of R in S^n and the homotopy classes of mappings of $\beta(R)$ in S^n .

b) We have still to show that the homomorphism of the cohomology groups induced by ψ^* is an isomorphism. Let \mathfrak{U} be any finite covering of R on whose nerve a cocycle Γ^n of R is defined, or on whose nerve a cohomology $\Gamma^n \sim \Gamma'^n$ holds. Let U_1, \dots, U_k be the sets of \mathfrak{U} . As in Theorem 1.1 we can define continuous functions (barycentric coordinates) f_i on R such that, if $x \in R - U_i$, $f_i(x) = 0$ and, if $x \in R$, $\sum f_i(x) = 1$. Let F_i be the continuous functions on $\beta(R)$ such that $f_i(x) = F_i\psi(x)$ for $x \in R$. Then $F_i(y) = 0$ if $y \in \psi(R - U_i)$ and hence by continuity also if $y \in \overline{\psi(R - U_i)}$. Also $\sum F_i(y) = 1$ if $y \in \psi R$, and hence by continuity also if $y \in \overline{\psi R} = \beta(R)$. Hence, if $y \in \beta(R)$, for some i , $F_i(y) \neq 0$ and $y \in \beta(R) - \overline{\psi(R - U_i)}$. Hence the open sets $V_i = \beta(R) - \overline{\psi(R - U_i)}$ form a covering \mathfrak{B} of $\beta(R)$. Hence the open sets $W_i = \psi^{-1}V_i$ form a covering \mathfrak{A} of R . Now $\psi(W_i) \subset V_i = \beta(R) - \overline{\psi(R - U_i)} \subset \beta(R) - \psi(R - U_i)$. Hence W_i and $R - U_i$ do not intersect. Hence $W_i \subset U_i$ and \mathfrak{A} is a refinement of \mathfrak{U} .

If any collection of open sets of \mathfrak{B} intersect in a point $p \in R$ then $\psi(p)$ is in each of the corresponding sets of \mathfrak{B} which therefore also intersect.

Conversely, if any collection of open sets of \mathfrak{B} intersect, their intersection contains points of ψR and hence the corresponding sets of \mathfrak{B} intersect. Hence the mapping of $N(\mathfrak{B})$ in $N(\mathfrak{B})$ induced by ψ is an isomorphism. Hence any cocycle or cohomology in the nerve of the refinement \mathfrak{B} of \mathfrak{U} is the image of a cocycle or cohomology in the nerve of a covering \mathfrak{B} of $\beta(R)$. Hence the mapping ψ^* induces an isomorphism of $H^n(\beta(R))$ on $H_F^n(R)$. This completes the proof.

COROLLARY 9.4. *Let R be a countably compact normal space with $\dim R \leq n$, ($n \geq 1$). The correspondence between the cocycle $f^*\Delta^n$ and the mapping f induces a 1—1 correspondence between the elements of the n -cohomology group $H_F^n(R)$ based on finite coverings and the classes of homotopic mappings of R into the n -sphere S^n .*

Proof. Since by Lemma 9.2 the classes of homotopic mappings of the countably compact space R are identical with the classes of uniformly homotopic mappings, the corollary follows immediately from Theorem 9.3.

THEOREM 9.5. *If R is a normal [countably compact normal] space, the correspondence between the cocycle $f^*\Delta^1$ and the mapping f induces an isomorphism between the 1-cohomology group $H_F^1(R)$ based on finite coverings and the group of uniform homotopy [homotopy] classes of mappings of R into the circumference S^1 .*

Proof. The proof of these forms of Bruschi's Theorem is adequately indicated by the proofs of Theorem 9.3 and Corollary 9.4.

Theorem 9.5 may be used to find the 1-dimensional cohomology groups, based on finite coverings, of spaces for which the group of uniform homotopy classes of mappings in S^1 can be calculated. It is difficult to find these cohomology groups directly, even in the simplest non-compact spaces. In Theorem 9.6 the 1-cohomology group of the straight line is described. The remarkable results indicate that the cohomology groups based on finite coverings are not suitable for describing the connectivity properties of non-compact spaces.

THEOREM 9.6. *The 1-cohomology group of the real line based on finite coverings is isomorphic with the difference group of the additive group of all real continuous functions on the real line modulo the group of all bounded continuous functions. The number of its elements is $\aleph = 2^{\aleph_0}$.*

Proof. Let R^1 be the real line and let f be any mapping of R^1 in the unit circle S^1 in the complex plane. Since every such mapping is homotopic to a constant mapping there is a continuous real function ϕ on R^1 such that,¹⁷ for $x \in R^1$, $f(x) = \exp(i\phi(x))$. To every such function ϕ corresponds a unique mapping f and to the sum or difference of such real functions ϕ corresponds the sum or difference of the mappings. Thus there is a homomorphism of the additive group of real continuous functions onto the group of mappings of R^1 in S^1 . The bounded real continuous functions, and only these,¹⁸ correspond to mappings of R^1 in S^1 , which are uniformly homotopic to constant mappings. Hence the difference group of the real continuous functions modulo the bounded functions is isomorphic with the group of uniform homotopy classes of mappings of R^1 in S^1 , which by Theorem 9.5 is isomorphic with the 1-cohomology group of R^1 based on finite coverings.

The power of the set of continuous real functions is \aleph . Also the set of functions $\phi(x) = kx$ with k real, which is in 1-1 correspondence with a subset of the set of elements of the above difference group, has power \aleph . Hence the number of elements in the 1-cohomology group of R^1 based on finite coverings is \aleph .

10. Applications to non-algebraic topology. We now apply some of the preceding methods and results to the proof of two theorems of non-algebraic topology. The first is a variation of Borsuk's Theorem in which uniform homotopy is replaced by ordinary homotopy. The second is a generalization of a theorem of Alexandroff ([2], p. 220) on transforming a given space to a cyclic space of the same dimension by contracting a closed set to a point.

The following lemma is needed in the proof of Borsuk's Theorem with ordinary homotopy.

LEMMA 10.1. *Let R' be a closed subset of a paracompact normal space R and let g be any continuous real function on $R' \times I$. Then for any $\epsilon > 0$ there is a continuous real function h on $R \times I$ such that, for all $(x, t) \in R' \times I$, $|h(x, t) - g(x, t)| < \epsilon$.*

Proof. If R satisfies the Hausdorff separation axiom then, according to a theorem of Dieudonné ([11], Theorem 5), $R \times I$ is a paracompact

¹⁷ See [13], p. 68.

¹⁸ See [13], p. 68.

Hausdorff space and hence normal. A continuous function g on a closed subset $R' \times I$ of the normal space $R \times I$ can then be extended over the normal space and the lemma follows.

Otherwise, let \mathfrak{B} be the covering of $R' \times I$ by the inverse images by g of a covering of the real line by intervals of length $< \epsilon$. Let $\mathfrak{B}_1 = \{V_{\alpha i}\}$ be a P -covering of $R' \times I$ which is a refinement of \mathfrak{B} (See proof of Lemma 6.3), and let $V_{\alpha i} = U'_\alpha \times W_{\alpha i}$ where $\mathfrak{U}' = \{U'_\alpha\}$ is a locally finite covering of R' . Let U'_α be extended to an open set U_α of R , such that $U'_\alpha = U_\alpha \cdot R'$, and let $U_0 = R - R'$. Then $\mathfrak{U} = \{U_\alpha\}$, $\alpha = 0, 1, \dots$, is a covering of R . Since R is paracompact and normal, there is a canonical mapping ϕ of R in the nerve N of \mathfrak{U} .

Let a point p_α be chosen in each set $U'_\alpha \in \mathfrak{U}'$. Let G_t be the mapping of N in the real line which maps each vertex u_α of N , $\alpha \neq 0$, in $g(p_\alpha, t)$ and maps u_0 in a fixed point, say the zero point, of the line, and maps N linearly. If $y \in N$, let $G(y, t) = G_t(y)$. Then G is continuous on each finite subcomplex of $N \times I$. If $x \in R$, let $h(x, t) = G(\phi(x), t)$. Then h is a continuous mapping of $R \times I$ in the real line.

Let $x \in R'$. Since \mathfrak{U}' is locally finite, x is in a finite number of open sets of \mathfrak{U}' and hence also is in a finite number of sets $U_\alpha, \dots, U_\gamma$ of \mathfrak{U} . Then ϕ maps x in the closure of the simplex $u_\alpha \cdots u_\gamma$ determined by x in N , a simplex which does not have u_0 as a vertex. If $x \in U_\alpha$, then $x \in U'_\alpha$ and hence also, for each i , $(x, t) \in U'_\alpha \times W_{\alpha i}$ for some i . Also $p_\alpha \in U'_\alpha$ and $(p_\alpha, t) \in U'_\alpha \times W_{\alpha i}$. Hence (x, t) and (p_α, t) are in the same set $V_{\alpha i}$ of the covering \mathfrak{B}_1 and hence also in the same set of the covering \mathfrak{B} . Hence $g(x, t)$ and $g(p_\alpha, t)$ are in the same interval of the ϵ -covering of the real line. Hence, since $G_t(u_\alpha) = g(p_\alpha, t)$, G_t maps each vertex, and hence each point, of the closed simplex determined by x in a point of the real line within ϵ of $g(x, t)$. But one of the points of this simplex is $\phi(x)$. Hence, since $G_t\phi(x) = h(x, t)$, $|h(x, t) - g(x, t)| < \epsilon$. This completes the proof of the lemma.

Now, for normal spaces which are paracompact, we are able to prove Borsuk's Theorem with ordinary homotopy replacing uniform homotopy.

THEOREM¹⁹ 10.2. *Let R' be a closed subset of a paracompact normal space R and let f_0 and f_1 be two homotopic mappings of R' into an absolute neighborhood retract S . If f_0 can be extended to a mapping F_0 of R into S , then f_1 can be similarly extended to F_1 with F_0 and F_1 homotopic.*

¹⁹ This theorem, for a more restricted class of spaces, appears in my dissertation, Princeton University (1938). The proof was first published in [20], p. 86. For compact spaces the theorem is equivalent to Borsuk's theorem.

Proof. We assume that S is a subset of the Hilbert cube I_ω . Let $\eta > 0$ be such that the closed η -neighborhood \bar{U} of S can be retracted onto S .

Let g be a mapping of $R' \times I$ in S such that, for $x \in R'$, $g(x, 0) = f_0(x)$, $g(x, 1) = f_1(x)$. Then it easily follows from Lemma 10.1 that there is a mapping h of $R \times I$ in I_ω such that, for $(x, t) \in R' \times I$, $\rho(g(x, t), h(x, t)) < \eta$. Then $h(R' \times I) \subset U$. Let $h_0(x, t)$ be the point dividing the segment $[F_0(x), h(x, 0)]$ in the ratio $t : 1 - t$. Then h_0 is a continuous mapping of $R \times I$ in I_ω such that $h_0(x, 0) = F_0(x)$, $h_0(x, 1) = h(x, 0)$, and such that $h_0(R' \times I) \subset U$. Since R is normal, the mapping f_1 of R' in I_ω can be extended to a mapping F'_1 of R in I_ω . Then there is a continuous mapping h_1 of $R \times I$ in I_ω such that $h_1(x, 0) = h(x, 1)$, $h_1(x, 1) = F'_1(x)$ and such that $h_1(R' \times I) \subset U$. Let h_2 be the continuous mapping of $R \times I$ in I_ω defined by $h_2(x, t) = h_0(x, 3t)$ if $0 \leq t \leq \frac{1}{3}$, $h_2(x, t) = h(x, 3t - 1)$ if $\frac{1}{3} \leq t \leq \frac{2}{3}$, and $h_2(x, t) = h_1(x, 3t - 2)$ if $\frac{2}{3} \leq t \leq 1$. Then $h_2(x, 0) = F_0(x)$, $h_2(x, 1) = F'_1(x)$, and $h_2(R' \times I) \subset U$.

Let $h_2^{-1}U = V$. Then V is a neighborhood of $R' \times I$ in $R \times I$. Since I is compact, every point $x \in R'$ has a neighborhood $W(x)$ in R such that $W(x) \times I \subset V$. Let W be the sum of all these neighborhoods. When W is an open set of R such that $R' \subset W$, $W \times I \subset V$ and hence $h_2(W \times I) \subset U$. Since R is normal there is a continuous real function $p(x)$ defined on R such that $0 \leq p(x) \leq 1$ and such that $p(x) = 1$ for $x \in R'$ and $p(x) = 0$ for $x \in R - W$. Let h_3 be the continuous mapping of $R \times I$ in I_ω defined by $h_3(x, t) = h_2(x, t \cdot p(x))$. Then $h_3(x, 0) = F_0(x)$, and, if $x \in R'$, $h_3(x, 1) = f_1(x)$. Also $h_3(R \times I) \subset U$. Let ϕ be a retraction of U on S and let $H = \phi h_3$. Then H is a mapping of $R \times I$ in S such that $H(x, 0) = F_0(x)$, and, if $x \in R'$, $H(x, 1) = f_1(x)$. Let $F_1(x) = H(x, 1)$. Then F_1 is an extension of f_1 homotopic to F_0 on S . This completes the proof of the theorem.

Let R' be a non-vacuous closed set of a space R . We say that a space R_1 is obtained from R by *contracting* R' to a point²⁰ if there is a continuous mapping ψ of R onto R_1 which maps R' into a closed set (p) consisting of a single point p of R_1 and which maps $R - R'$ homeomorphically onto $R_1 - (p)$.

If R is paracompact, R_1 is also paracompact. For let $\mathcal{U} = \{U_\alpha\}$ be any covering of R_1 by open sets. Then $\{\psi^{-1}U_\alpha\}$ is a covering of R which, since R is paracompact, has a locally finite refinement $\mathcal{V} = \{V_\beta\}$. Then $\{(R - R') \cdot V_\beta\}$ is a locally finite covering of $R - R'$ which, by the homeomorphism $\psi|_{R - R'}$, is transformed into a locally finite covering $\{\psi((R - R') \cdot V_\beta)\}$ of $R_1 - (p)$ by open sets. By including one open set

²⁰ See [2], p. 220.

of \mathcal{U} containing p , we get a locally finite covering of R_1 which is a refinement of \mathcal{U} . Hence R_1 is paracompact.

Similarly if R is compact R_1 is also compact. If R has a countable base for its open sets and if p has a countable base for its neighborhoods in R_1 then R_1 has a countable base for its open sets. For, since R has a countable base, $R - R'$ also has a countable base, and hence also $R_1 - (p)$ has a countable base. This together with the countable base for the neighborhoods of p forms a countable base for R_1 .

If R has dimension $\leq n$, and if every open set of R_1 containing p contains the closure of an open set containing p then R_1 has dimension $\leq n$. For let $\mathcal{U} = \{U_\alpha\}$ be any covering of R_1 . If $p \in U_\gamma \in \mathcal{U}$, let G_1, G_2 be neighborhoods of p such that $\bar{G}_1 \subset G_2, \bar{G}_2 \subset U_\gamma$. Then U_γ together with all the open sets $U_\alpha - \bar{G}_2$ form a covering \mathcal{U}_1 of R_1 . Then $\psi^{-1}\mathcal{U}_1$ is a covering of R which has a refinement \mathfrak{B}_1 of order $\leq n + 1$. All sets of \mathfrak{B}_1 which intersect $\psi^{-1}\bar{G}_2$ are contained in $\psi^{-1}U_\gamma \in \mathfrak{B}_1$. These sets may be replaced by a single set, their sum, which we call V_0 , without increasing the order of the covering, and the resultant covering \mathfrak{B}_2 is still a refinement of $\psi^{-1}\mathcal{U}_1$. Then all sets of \mathfrak{B}_2 , except V_0 , are open sets of $R - R'$ and their images by ψ are open sets of R_1 . The set V_0 is the sum of the open sets $\psi^{-1}G_2$ and $V_0 - \psi^{-1}\bar{G}_1$. The image by ψ of $\psi^{-1}G_2$ is the open set G_2 of R_1 and the image of the open set $V_0 - \psi^{-1}\bar{G}_1$ of $R - R'$ is an open set of $R_1 - (p)$. Hence ψV_0 is an open set of R_1 . Hence ψ maps the covering \mathfrak{B}_2 into an open covering \mathcal{U}_2 of R_1 . The order of \mathcal{U}_2 is the same as that of \mathfrak{B}_2 and since \mathfrak{B}_2 is a refinement of \mathfrak{B}_1 , \mathcal{U}_2 is a refinement of \mathcal{U}_1 and hence of \mathcal{U} . Hence \mathcal{U} has a refinement of order $\leq n + 1$. Hence R_1 has dimension $\leq n$.

If R is normal and if every open set of R_1 containing p contains the closure of an open set containing p , then R_1 is also normal. For let F_1 and F_2 be any two non-intersecting closed sets of R_1 . The point p cannot be in both closed sets. Suppose that p is not in F_2 . Let G_1 and G_2 be neighborhoods of p such that $\bar{G}_1 \subset G_2$ and $\bar{G}_2 \subset R_1 - F_2$. Then $R_1 - G_1$ is homeomorphic to a closed set of R and hence is normal. Hence there are non-intersecting open sets V_1 and V_2 of $R_1 - G_1$ containing the closed sets $(R_1 - G_1) \cdot F_1$ and F_2 respectively. Then $W_1 = G_2 + (R_1 - \bar{G}_1) \cdot V_1$ and $W_2 = (R - \bar{G}_2) \cdot V_2$ are non-intersecting open sets of R_1 which contain F_1 and F_2 respectively. Hence R_1 is normal.

An n -dimensional space is called cyclic if there is an essential mapping of R on the n -sphere S^n . The theorem of Alexandroff ([2] p. 220)²¹ on

²¹ See also [19], p. 195

obtaining a cyclic space by contracting a closed set to a point can be stated as follows for paracompact normal spaces.

THEOREM 10.3. *If R is a paracompact normal space of dimension n , ($n \leq 1$), then, by contracting a suitable closed set R' of R to a point, R is transformed into a paracompact normal cyclic space R_1 of dimension n .*

Proof. By Corollary 3.5 there is a closed set R' of R and a mapping f of R' in the $(n-1)$ -sphere S^{n-1} which cannot be extended to a mapping of R in S^{n-1} . Let S^{n-1} be the boundary of a simplex σ^n and let F be an extension of f to a mapping of R in σ^n . Then $R' \subset F^{-1}S^{n-1}$. Let R' be replaced by a larger set, if necessary, so that $R' = F^{-1}S^{n-1}$. Let χ be a simplicial mapping of a subdivision of σ^n onto a sphere S^n such that S^{n-1} is mapped into a point q of S^n , with $\chi^{-1}(q) = S^{n-1}$, and such that the degree of the mapping is one.

We define a space R_1 , a mapping ψ of R on R_1 and a mapping g of R_1 on S^n as follows: The space R_1 consists of a set of points in 1-1 correspondence with the points of $R - R'$ and one additional point p . The mapping ψ maps each point of $R - R'$ on its corresponding point of R and maps R' into p . The mapping g maps each point $x \in R_1 - (p)$ into $\chi F \psi^{-1}(x)$ and maps p into q . Thus, for each point $y \in R$, $g\psi(y) = \chi F(y)$. Let the open sets of R_1 be the images by ψ of the open sets of $R - R'$ and the counter images by g of the open sets of S . Thus the mappings ψ and g are continuous, the set (p) is closed, and the partial mapping $\psi|_{R-R'}$ is a homeomorphism. Hence R_1 is a space obtained from R by contracting R' to a point. The neighborhoods of q with radius $1/r$, where r is an integer, form a basis for the neighborhoods of q such that each contains the closure of another. The inverse images by g of these neighborhoods form a similar basis for the neighborhood of p in R . It follows that R_1 is a paracompact normal space of dimension $\leq n$ which is compact if R is compact and has a countable base if R has a countable base. Since also the intersection of the neighborhoods of p consists of p alone, R_1 can be shown to satisfy the T_0 , T_1 , T_2 , or T_3 separation axioms if R does.

In proving the theorem we first take the case $n \geq 2$. The fundamental cocycle Δ^n of S^n is cohomologous in S^n to a cocycle of $S^n - (q)$. Replacing Δ^n , if necessary, by this cocycle we assume that Δ^n is a cocycle of $S^n - (q)$. Then $\chi^*\Delta^n$ is a cocycle of $\sigma^n - S^{n-1}$. Since the mapping χ is of degree one, $\chi^*\Delta^n \sim \pm \Psi\Delta^{n-1}$ where $\Psi\Delta^{n-1}$ is the coboundary in $\sigma^n - S^{n-1}$ of the fundamental cocycle Δ^{n-1} of S^{n-1} . We assume that $\chi^*\Delta^n \sim \Psi\Delta^{n-1}$. Then $F^*\chi^*\Delta^n \sim F^*\Psi\Delta^{n-1} \sim \Psi f^*\Delta^{n-1}$ in $R - R'$. But (See Theorem 5.2) $\Psi f^*\Delta^{n-1}$ and

hence also $F^*\chi^*\Delta^n$ is not ~ 0 in $R - R'$. Since $g\psi = \chi F$, $\psi^*g^*\Delta^n \sim F^*\chi^*\Delta^n$ in $R - R'$. Hence $\psi^*g^*\Delta^n$ is not ~ 0 in $R - R'$. Hence $g^*\Delta^n$ is not ~ 0 in $R_1 - (p)$. But every covering of R_1 has a refinement such that p is in only one open set of the refinement. Hence, since $n \geq 2$, $g^*\Delta^n$ is not ~ 0 in R_1 . It follows that the dimension of R_1 is at least n , and hence is exactly n . It also follows, by Theorem 7.5, that g is an essential mapping of R_1 on S^n . Hence R_1 is a cyclic space of dimension n .

Now let $n = 1$. We assume that σ^1 is the line interval $(0, 1)$ and hence S^0 consist of the points 0 and 1. We also suppose that S^1 is the unit circle in the complex plane, that q is the unit point, and that $\chi(t) = \exp(2\pi it)$ for $t \in \sigma^1$. Suppose, if possible, that g is an inessential mapping of R_1 in S^1 . Then, according to a theorem of Eilenberg ([13], p. 68) there is a real function ϕ of the points of R_1 such that $g(y) = \exp(2\pi i\phi(y))$ for $y \in R_1$. Since $g(y) = q$ only if $y = p$, $\phi(y)$ is an integer only if $y = p$. We may assume without loss of generality that $\phi(p) = 0$. Since $g\psi = \chi F$, if $x \in R$, $\exp(2\pi iF(x)) = \exp(2\pi i\phi\psi(x))$ and hence $\phi\psi(x) = F(x) \bmod 1$. Let A be the subset of R consisting of $F^{-1}(0) + \psi^{-1}\phi^{-1}(0, 1)$ where $(0, 1)$ is the open interval from 0 to 1. Using square brackets for closed intervals we can write, since $\phi\psi(x) = F(x) \bmod 1$, $A = F^{-1}[0, \frac{1}{2}] \cdot \psi^{-1}\phi^{-1}[0, \frac{1}{2}] + \psi^{-1}\phi^{-1}[\frac{1}{2}, 1]$. Thus A is a closed set. Also $A = F^{-1}[0, \frac{1}{2}] \cdot \psi^{-1}\phi^{-1}(-\frac{1}{4}, \frac{1}{2}) + \psi^{-1}\phi^{-1}(\frac{1}{4}, 1)$. Thus A is also an open set. Hence $B = R - A$ is a closed set. Now $f^{-1}(0) = F^{-1}(0) \subset A$ and $f^{-1}(1) = F^{-1}(1) \subset B$. Let F_1 be the mapping of R in S^0 which maps A in 0 and B in 1. Then F_1 is a continuous extension of f contrary to the assumption that f has no extension. Hence g is an essential mapping of R_1 in S^1 . It follows that R_1 has dimension at least 1, and hence exactly 1. Hence R_1 is a cyclic space of dimension 1. This completes the proof of the theorem.

It may be noted that in the 1-dimensional case we have not used the assumption of paracompactness. It follows that any 1-dimensional normal space can be transformed, by contracting a closed set to a point, into a 1-dimensional normal cyclic space.

If in the proof of this theorem we used finite coverings and cocycles based on finite coverings and if we used Theorem 9.3 instead of Theorem 7.5 we could have shown that any n -dimensional normal space can be transformed, by contracting a closed set to a point, into an n -dimensional normal space which can be mapped essentially, in the sense of uniform homotopy, on the n -sphere.

11. Compact cocycles. It was shown by Borsuk and Eilenberg [6, 14] that the group of homotopy classes of mappings of the complement R of a

compact set of Euclidean space into the circumference S^1 is isomorphic with the cohomology group of R based on compact cocycles. This cohomology group is defined as follows:

Let $C_1, C_2, \dots, C_a, \dots$, be the compact subsets of a space R . A compact n -cocycle of R is a set $\{\Gamma_a^n\}$ of cocycles, one in each of the compact subsets $\{C_a\}$ such that, if $C_\beta \subset C_a$ and if ψ_a^β is the identical mapping of C_β in C_a , then $\psi_a^{\beta*} \Gamma_a^n \sim \Gamma_\beta^n$ in C_β . Two compact n -cocycles $\{\Gamma_a^n\}$ and $\{\Gamma'_a^n\}$ are cohomologous if $\Gamma_a^n \sim \Gamma'_a^n$ in C_a for every C_a . Also $\{\Gamma_a^n\} \sim \{\Gamma'_a^n\} - \{\Gamma''_a^n\}$ means $\Gamma_a^n \sim \Gamma'_a^n - \Gamma''_a^n$ in C_a for every C_a . The cohomology group of R based on compact cocycles is the group of classes of cohomologous compact cocycles of R .

If $\{D_\gamma\}$ is any subfamily of the compact subsets of R , a set $\{\Gamma_\gamma^n\}$ of cocycles, one on each of the sets $\{D_\gamma\}$, is called a cocycle of the subfamily if, whenever $D_\delta \subset D_\gamma$, if ψ_γ^δ is the identical mapping of D_δ in D_γ , $\psi_\gamma^{\delta*} \Gamma_\gamma^n \sim \Gamma_\delta^n$ in D_δ . Similarly one defines the sum of two cocycles of the family or cohomology of two cocycles of the family. Then the classes of cohomologous cocycles of the family form the cohomology group of the family. A family $\{D_\gamma\}$ of compact subsets of R is called a cofinal family if every compact set of R is a subset of one of the sets $\{D_\gamma\}$. A cocycle of a cofinal family of compact sets determines a unique compact cocycle of R and the cohomology group of a complete family of compact sets is isomorphic with the cohomology group of R based on compact cocycles.

If Γ^n is an n -cocycle of R and if, for each compact set C_a of R , χ_a is the identical mapping of C_a in R , then $\chi_a^* \Gamma^n$ is an n -cocycle of C_a and $\{\chi_a^* \Gamma^n\}$ is a compact n -cocycle of R . If f is a mapping of R into S and if Γ^n is an n -cocycle of S , then $f^* \Gamma^n$ is a cocycle of R and $\{\chi_a^* f^* \Gamma^n\}$ is a compact cocycle of R . If $\Gamma^n \sim \Gamma'^n$ in S , then $\chi_a^* f^* \Gamma^n \sim \chi_a^* f^* \Gamma'^n$ in C_a , and $\{\chi_a^* f^* \Gamma^n\} \sim \{\chi_a^* f^* \Gamma'^n\}$ in R . If $\Gamma^n \sim \Gamma'^n - \Gamma''^n$ in S , then $\chi_a^* f^* \Gamma^n \sim \chi_a^* f^* \Gamma'^n - \chi_a^* f^* \Gamma''^n$ in C_a , and $\{\chi_a^* f^* \Gamma^n\} \sim \{\chi_a^* f^* \Gamma'^n\} - \{\chi_a^* f^* \Gamma''^n\}$ in R . Thus f induces a homomorphism of the n -cohomology group of S into the n -cohomology group of R based on compact cocycles.

Let R be a paracompact normal space which is locally compact and locally connected. The theorem of Borsuk and Eilenberg will be shown to hold for such a space.

Since R is locally connected, each component of R is an open set. For, if x is any point of R , any neighborhood of x contains a neighborhood which is contained in the component of R containing x . Hence the components V_a of R form an open covering $\mathfrak{B} = \{V_a\}$. Since R is locally compact, each point of R has a neighborhood whose closure is compact. Hence R has a covering

$\mathfrak{B} = \{W_\alpha\}$ consisting of all open sets W_α with compact closures. Since R is paracompact, for any covering \mathfrak{U} of R there is a common refinement \mathfrak{U}_1 of \mathfrak{U} , \mathfrak{B} , and \mathfrak{B} , which is locally finite. Since R is normal, we may further assume that \mathfrak{U}_1 is a normal covering of R , i. e., that there is a canonical mapping ϕ of R on the nerve N_1 of \mathfrak{U}_1 which is essential on every simplex of N_1 . The star of each vertex u of N_1 is the image of an open set U of \mathfrak{U}_1 , which is contained in an open set of \mathfrak{B} whose closure is compact. Hence \bar{U} is compact and its image, which is the closure of the star of u , is compact and hence is a finite complex. Thus \mathfrak{U}_1 is a star-finite covering and R is an s -space.

Since \mathfrak{U}_1 is a refinement of the covering \mathfrak{B} by components, it follows that if U_α and U_β are sets of \mathfrak{U}_1 contained in distinct sets of \mathfrak{B} , i. e., in distinct components, the vertices u_α and u_β of N_1 are not connected. Thus ϕ does not map more than one component of R in any component of N_1 . But the components of the complex N_1 are open and closed. Hence, since ϕ is continuous, the inverse image of a component of N_1 is an open and closed set of a component of R , and hence is a component of R .

Now, if L_1 is the star of a vertex of N_1 , $A = \overline{\phi^{-1}(L_1)}$ is compact. Let L_2 be the star of the closed complex \bar{L}_1 . Then $A_2 = \overline{\phi^{-1}(L_2)}$ is the sum of a finite number of compact sets and is therefore compact. Also A_1 is contained in the interior of A_2 ; $A_1 \subseteq A_2$. Proceeding thus we get a sequence $\{L_i\}$ where L_{i+1} is the star of the finite complex \bar{L}_i and a sequence of compact sets $\{A_i\}$ where $A_i = \overline{\phi^{-1}(L_i)}$ and $A_i \subseteq A_{i+1}$. The sum ΣL_i of the sequence is seen to be a component of N_1 . Hence ΣA_i is a component of R . Let p be a point of A_1 . Let B_i be the component of A_i containing p . Then B_i is closed and therefore compact. Since R is locally connected, B_i is contained in an open set contained in a connected set of the interior of A_{i+1} . Hence $B_i \subseteq B_{i+1}$. Also $\Sigma B_i = \Sigma A_i$. For otherwise there would be a point x of the connected set ΣA_i such that every neighborhood of x contains a point of ΣB_i and a point of $\Sigma A_i - \Sigma B_i$. But x is in the interior of some set A_i and has a neighborhood contained in a connected set of the interior of A_i . This neighborhood contains a point of some set B_i and a point of $A_i - \Sigma B_k$. Since the sets A_i are increasing it may be assumed that $i > j$. But then B_i would contain the whole neighborhood of x . Hence $\Sigma B_i = \Sigma A_i$. Thus each component of R is the sum of a sequence of compact connected sets, each being contained in the interior of the next set of the sequence.

LEMMA 11.1. *If $\{\Gamma_\alpha^1\}$ is a compact 1-cocycle of the locally connected locally compact paracompact normal space R there is a mapping f of R into the circumference S^1 such that $\{\chi_\alpha^* f^* \Delta^1\} \sim \{\Gamma_\alpha^1\}$ in R .*

Proof. Let $V_\beta \in \mathfrak{B}$ be a component of R and let $\{B_{\beta i}\}$ be a sequence of

compact connected sets such that $B_{\beta i} \subseteq B_{\beta, i+1}$ and $V_\beta = \sum_i B_{\beta i}$. Then, by Theorem 8.1, there is a mapping $f_{\beta 1}$ of $B_{\beta 1}$ into S^1 such that $f_{\beta 1}^* \Delta^1 \sim \Gamma_{\beta 1}^1$ in $B_{\beta 1}$, where $\Gamma_{\beta 1}^1$ is the cocycle of $\{\Gamma_a^1\}$ in the compact set $B_{\beta 1}$. Assume we have a mapping $f_{\beta i}$ of $B_{\beta i}$ into S^1 such that $f_{\beta i}^* \Delta^1 \sim \Gamma_{\beta i}^1$ in $B_{\beta i}$. There exists a mapping $g_{\beta, i+1}$ of $B_{\beta, i+1}$ into S^1 such that $g_{\beta, i+1}^* \Delta^1 \sim \Gamma_{\beta, i+1}^1$ in $B_{\beta, i+1}$. If ψ_i is the identical mapping of $B_{\beta i}$ in $B_{\beta, i+1}$, then $\psi_i^* g_{\beta, i+1}^* \Delta^1 \sim \psi_i^* \Gamma_{\beta, i+1}^1 \sim \Gamma_{\beta i}^1 \sim f_{\beta i}^* \Delta^1$ in $B_{\beta i}$. Hence $g_{\beta, i+1} \psi_i = g_{\beta, i+1} | B_{\beta i}$ is homotopic to $f_{\beta i}$. Hence, by Theorem 10.2, $f_{\beta i}$ can be extended to a mapping $f_{\beta, i+1}$ of $B_{\beta, i+1}$ in S^1 , with $f_{\beta, i+1}$ homotopic to $g_{\beta, i+1}$. Hence $f_{\beta, i+1}^* \Delta^1 \sim g_{\beta, i+1}^* \Delta^1 \sim \Gamma_{\beta, i+1}^1$ in $B_{\beta, i+1}$. Thus there is a sequence of mappings $f_{\beta i}$ of the sets $B_{\beta i}$ in S^1 such that $f_{\beta i}^* \Delta^1 \sim \Gamma_{\beta i}^1$ in $B_{\beta i}$ and such that $f_{\beta, i+1}$ is an extension of $f_{\beta i}$. Since $B_{\beta i} \subseteq B_{\beta, i+1}$, the limit mapping f_β of V_β in S^1 is continuous. The mapping f of R in S^1 which coincides with f_β in each V_β is the required mapping.

Let C_a be a set consisting of a finite sum of sets $B_{\beta i}$, $B_{\gamma j}$, etc. each from a different component of R . Then there is a cocycle Γ_a^1 of C_a uniquely determined up to cohomology such that for each of the sets $B_{\beta i}$, if $\psi_a^{\beta i}$ is the identical mapping of $B_{\beta i}$ in C_a , $\psi_a^{\beta i} \Gamma_a^1 \sim \Gamma_{\beta i}^1$ in $B_{\beta i}$. The cocycle Γ_a^1 can be found by formally adding the cocycles $\Gamma_{\beta i}^1$ of the sets $B_{\beta i}$ in the nerve of each covering which separates these sets. Since Γ_a^1 is unique, $\Gamma_a^1 \sim \Gamma_a^1$ in C_a and $\Gamma_a^1 \sim \chi_a^* f^* \Delta^1$ in C_a and hence $\Gamma_a^1 \sim \chi_a^* f^* \Delta^1$ in C_a . Now the interiors of the sets $B_{\beta i}$ form a covering of R by open sets and hence any compact set is contained in a finite number of these. Hence any compact set of R is contained in a set C_a consisting of a finite sum of sets $B_{\beta i}$. Hence the sets C_a form a cofinal family of compact sets and thus $\Gamma_a^1 \sim \chi_a^* f^* \Delta^1$ in each of the compact sets C_a of a cofinal family. Hence $\{\Gamma_a^1\} \sim \{\chi_a^* f^* \Delta^1\}$. This completes the proof of the lemma.

LEMMA 11.2. *If f is a mapping of the locally connected locally compact paracompact normal space R into the circumference S^1 such that $\{\chi_a^* f^* \Delta^1\} \sim 0$ in R , then f is inessential.*

Proof. Since $\{\chi_a^* f^* \Delta^1\} \sim 0$ in R , $\chi_a^* f^* \Delta^1 \sim 0$ in C_a , and hence for each compact set C_a of R the partial mapping $f | C_a$ is inessential. In particular the partial mappings $f | B_{\beta i}$ of the sets $B_{\beta i}$ into S^1 are inessential. Hence there is a real function $\phi_{\beta i}$ on $B_{\beta i}$ such that for $x \in B_{\beta i}$, $f(x) = \exp(2\pi i \phi_{\beta i}(x))$. If a point x_β is chosen in $B_{\beta 1}$, $\phi_{\beta i}$ can be chosen so that $0 \leq \phi_{\beta i}(x_\beta) < 1$. Then, since $B_{\beta i}$ is connected, $\phi_{\beta i}$ is uniquely determined.²² Hence $\phi_{\beta, i+1}$ is an extension of $\phi_{\beta i}$. Hence there is a limit function ϕ_β defined and continuous on V_β such that for $x \in V_\beta$, $f(x) = \exp(2\pi i \phi_\beta(x))$. The function ϕ which

²² See [13], p. 64

coincides with ϕ_β on V_β for each component V_β is then continuous and $f(x) = \exp(2\pi i \phi(x))$ for $x \in R$. Hence the mapping f of R in S^1 is inessential.

THEOREM 11.3. *If a paracompact normal space R is locally compact and locally connected, the group of homotopy classes of mappings of R into the circumference S^1 is isomorphic with the 1-cohomology group of R based on compact cocycles.*

Proof. To each mapping f of R into S^1 corresponds a compact cocycle $\{\chi^*_{af^*\Delta^1}\}$. If f_0 and f_1 are homotopic mappings of R into S^1 , $f_0|C_a$ and $f_1|C_a$ are homotopic mappings of C_a in S^1 , and hence $\chi^*_{af^*_0\Delta^1} \sim \chi^*_{af^*_1\Delta^1}$ in C_a , and hence $\{\chi^*_{af^*_0\Delta^1}\} \sim \{\chi^*_{af^*_1\Delta^1}\}$ in R . If f is the difference of the two mappings f_0 and f_1 of R into S^1 , $f|C_a = f_0|C_a - f_1|C_a$, and hence $\chi^*_{af^*\Delta^1} \sim \chi^*_{af^*_0\Delta^1} - \chi^*_{af^*_1\Delta^1}$ in C_a , and hence $\{\chi^*_{af^*\Delta^1}\} \sim \{\chi^*_{af^*_0\Delta^1}\} - \{\chi^*_{af^*_1\Delta^1}\}$ in R . Thus there is a homomorphism of the group of homotopy classes into the cohomology group. But, by Lemma 11.1, there is a mapping for every cocycle. Hence the homomorphism is onto the cohomology group. By Lemma 11.2, only the class of inessential mappings corresponds to the zero element of the cohomology group. Hence the homomorphism is an isomorphism. This completes the proof.

It is easily seen by counterexamples that the requirements that R be locally connected and locally compact are essential for Theorem 11.3. And even if these conditions are satisfied the n -cohomology group of R need not be isomorphic with the n -cohomology group based on compact cocycles for $n > 1$. This may be shown by the following example.

Let R be the space defined as follows: Let the end circles of a sequence of cylindrical surfaces be distinguished as right and left. Let each point of the right circle of the cylinder R_i be identified with its diametrically opposite point and let the circle so obtained be identified with the left circle of R_{i+1} .

The space R can be subdivided to form an infinite complex K . Such subdivisions are obtained by subdividing each of the cylinders R_i so that the subdivisions agree on the common circles. If K_a is any such complex the covering \mathcal{U}_a of R by the stars of the vertices of K_a has K_a as nerve. Such coverings \mathcal{U}_a form a cofinal family of coverings of R . Let σ^2_{ai} be a designated 2-simplex of K_a contained in R_i and let σ^1_{ai} be a designated 1-simplex of K_a in the left circle of R_i . Then each 2-cocycle of K_a is cohomologous to a cocycle $c_a^i \sigma^2_{ai}$, and it is cohomologous to zero if and only if there exist numbers a_a^i such that $c_a^i \sigma^2_{ai} = a_a^i C^2_{ai}$ where, if $r > 1$, $C^2_{ai} = -2\sigma^2_{a,i-1} + \sigma^2_{ai}$, and $C^2_{a1} = \sigma^2_{a1}$. We assume that K_a is oriented so that C^2_{a1} is cohomologous to the coboundary of σ^1_{a1} .

Let Γ_a^2 be the 2-cocycle $c_a^i \sigma_{a^i}^2$ of K_a , where $c_a^i = 1$ or 0 according as i is odd or even. Then $\Gamma^2 = \{\Gamma_a^2\}$ is a cocycle of the cofinal family of coverings. Assume if possible that $\Gamma^2 \sim 0$ in R . Then $\Gamma_a^2 \sim 0$ in K_a for some covering \mathfrak{U}_a of the complete family. Then there are numbers a_a^i such that

$$\Gamma_a^2 = a_a^i C_{a^i}^2 = a_a^1 \sigma_{a^1}^2 + \sum_{i=2}^{\infty} (-2a_a^i \sigma_{a^i, i-1}^2 + a_a^i \sigma_{a^i}^2) = \sum_{i=1}^{\infty} (-2a_a^{i+1} + a_a^i) \sigma_{a^i}^2.$$

Hence

$$c_a^i = -2a_a^{i+1} + a_a^i, \text{ i. e., } -2a_a^{2n+1} + a_a^{2n} = 0, -2a_a^{2n+2} + a_a^{2n+1} = 1.$$

Therefore

$$\begin{aligned} 2^{2n+1} a_a^{2n+2} &= 2^{2n-1} a_a^{2n} - 2^{2n} = 2a_a^2 - \sum_{p=0}^{n-1} 2^{2(n-p)} \\ &= a_a^1 - \sum_{p=0}^n 2^{2(n-p)} = a_a^1 - \frac{1}{3}(2^{2n+2} - 1). \end{aligned}$$

Therefore

$$a_a^1 \equiv \frac{1}{3}(2^{2n+2} - 1) \pmod{2^{2n+1}}.$$

Therefore

$$a_a^1 \equiv \frac{1}{3}(2^{2n+2} - 1) - 2^{2n+1} = -\frac{1}{3}(2^{2n+1} + 1) \pmod{2^{2n+1}}.$$

Therefore

$$|a_a^1| \geq \frac{1}{3}(2^{2n+1} - 1) \geq 2^{2n-1} \text{ for every } n.$$

But this is impossible. Hence Γ^2 is not ~ 0 in R .

Therefore not all 2-cocycles of R are cohomologous to zero in R and hence R has essential mappings on S^1 . However in a compact subset of R consisting of a finite number of the elementary cylinders (with indentifications) all 2-cocycles are ~ 0 . But these compact subsets form a cofinal family of compact subsets of R . Hence every compact 2-cocycle of R is ~ 0 . Therefore the 2-cohomology group of R based on infinite coverings is not isomorphic with the one based on compact cocycles.

TUFTS COLLEGE,
MEDFORD, MASS.

BIBLIOGRAPHY

1. P. Alexandroff, "Untersuchungen über Gestalt und Lage abgeschlossener Mengen beliebiger Dimension," *Annals of Mathematics* (2), vol. 30 (1929), pp. 101-187.
2. P. Alexandroff, "Dimensionstheorie," *Mathematische Annalen*, vol. 106 (1932), pp. 161-238.
3. P. Alexandroff and H. Hopf, *Topologie*, Berlin (1935)

4. K. Borsuk, "Über eine Klasse von lokal zusammenhängenden Räumen," *Fundamenta Mathematicae*, vol. 19 (1932), pp. 220-242.
5. K. Borsuk, "Sur les prolongements des transformations continues," *Fundamenta Mathematicae*, vol. 28 (1937), pp. 99-110.
6. K. Borsuk and S. Eilenberg, "Über stetige Abbildungen der Teilmengen euklidischer Räume auf die Kreislinie," *Fundamenta Mathematicae*, vol. 26 (1936), pp. 207-223.
7. N. Brusilinsky, "Stetige Abbildungen und Bettische Gruppen der Dimensionszahlen 1 und 3," *Mathematische Annalen*, vol. 109 (1934), pp. 525-537.
8. E. Čech, "Théorie générale de l'homologie dans un espace quelconque," *Fundamenta Mathematicae*, vol. 19 (1932), pp. 149-183.
9. E. Čech, "Sur la connexité locale d'ordre supérieur," *Compositio Mathematica*, vol. 2 (1935), pp. 1-25.
10. E. Čech, "On bicomact spaces," *Annals of Mathematics* (2), vol. 38 (1937), pp. 823-844.
11. J. Dieudonné, "Une généralisation des espaces compacts," *Journal de Mathématiques pures et appliquées*, vol. 23 (1944), pp. 65-76.
12. C. H. Dowker, "Hopf's theorem for non-compact spaces," *Proceedings of the National Academy of Sciences*, vol. 23 (1937), pp. 293-294.
13. S. Eilenberg, "Transformations continues en circonférence et la topologie du plan," *Fundamenta Mathematicae*, vol. 26 (1936), pp. 61-112.
14. S. Eilenberg, "Bemerkungen zur Pontrjaginschen Verallgemeinerung des Alexander'schen Dualitätssatzes," *Fundamenta Mathematicae*, vol. 26 (1936), pp. 224-228.
15. S. Eilenberg, "Extension and classification of continuous mappings," *Lectures in Topology*, Ann Arbor (1941), pp. 57-99.
16. H. Freudenthal, "Die Hopf'sche Gruppe," *Compositio Mathematica*, vol. 2 (1935), pp. 134-162.
17. H. Hopf, "Die Klassen der Abbildungen der n -dimensionalen Polyeder auf die n -dimensionale Sphäre," *Commentarii Mathematici Helvetici*, vol. 5 (1933), pp. 39-54.
18. W. Hurewicz, "Über Abbildungen von endlichdimensionalen Räumen auf Teilmengen Cartesischen Räume," *Sitzungsberichte der Preussischen Akademie der Wissenschaften* (1933), pp. 754-768.
19. W. Hurewicz, "Sur la dimension des produits Cartésiens," *Annals of Mathematics* (2), vol. 36 (1935), pp. 194-197.
20. W. Hurewicz and H. Wallman, *Dimension Theory*, Princeton (1941).
21. S. Kaplan, "Homology properties of arbitrary subsets of Euclidean spaces," to appear in the *Transactions of the American Mathematical Society*.
22. C. Kuratowski, "Sur un théorème fondamental concernant le nerf d'un système d'ensembles," *Fundamenta Mathematicae*, vol. 20 (1933), pp. 191-196.
23. S. Lefschetz, "Abstract complexes," *Lectures in Topology*, Ann Arbor (1941), pp. 1-28.
24. S. Lefschetz, *Algebraic Topology*, New York (1942).
25. S. Lefschetz, *Topics in Topology*, Princeton (1942).
26. A. Weil, "Sur les espaces à structure uniforme et sur la topologie générale" (*Actualités Scientifiques et Industrielles*, No. 551), Paris (1938).
27. H. Whitney, "On matrices of integers and combinatorial topology," *Duke Mathematical Journal*, vol. 3 (1937), pp. 35-45.
28. H. Whitney, "The maps of an n -complex into an n -sphere," *Duke Mathematical Journal*, vol. 3 (1937), pp. 51-55.

ON THE ZETA-FUNCTIONS OF ALGEBRAIC NUMBER FIELDS.*

By RICHARD BRAUER.

1. It was proved by E. Artin¹ that if k is an algebraic number field (of finite degree) and K a normal extension field with the icosahedral group as the Galois group with regard to k , then the zeta-function $\zeta(s, k)$ of k divides the zeta-function $\zeta(s, K)$, in the sense that the quotient $\zeta(s, K)/\zeta(s, k)$ is an integral function of the complex variable s . Using Artin's method, we shall show in this note that *the zeta-function $\zeta(s, k)$ of an algebraic number field k divides the zeta-function $\zeta(s, K)$ of every normal extension field K of k .*^{1a} The proof (3) is based on a group-theoretical lemma which is established in 2. Our result will enable us to prove in 4 the following theorem conjectured by C. L. Siegel.² *Consider all algebraic number fields of a fixed degree n . Let d be the discriminant of k , let h be the number of classes of ideals in k , and let R be the regulator of k . Then*

$$(*) \quad \log(hR) \sim \log \sqrt{|d|}, \quad (\text{for } |d| \rightarrow \infty).$$

This had been proved by Siegel for quadratic fields k . In the case of imaginary quadratic fields k , the result gives a refinement of the famous theorem of Heilbronn that h tends to infinity with $|d|$. Siegel also mentions that his method makes it possible to establish (*) for all fields k of fixed degree n which are *soluble* with regard to a fixed subfield. Using the above results on the zeta-function, and some results of Landau, we can apply Siegel's original method to obtain (*) in its full generality.

2. Let G be a group of finite order g . The irreducible characters of G will be denoted by $\chi_0, \chi_1, \dots, \chi_{n-1}$ where in particular χ_0 is the 1-character, $\chi_0(\sigma) = 1$ for all σ in G . If f_ν is the degree of χ_ν , the character θ of the regular representation of G is given by

* Received January 9, 1947; Presented to the American Mathematical Society, April 25, 1947.

¹ E. Artin, *Mathematische Annalen*, vol. 89 (1923), pp. 147-156; *Abhandlungen aus dem Mathematischen Seminar Hamburg*, vol. 3 (1924), pp. 89-108. Cf. also E. Artin, *Abhandlungen aus dem Mathematischen Seminar Hamburg*, vol. 8 (1931), pp. 292-306.

^{1a} After I had submitted the present paper, H. W. Brinkmann kindly drew my attention to the fact that this Theorem 1 had already been obtained by Hideo Aramata, *Proceedings of the Imperial Academy of Japan*, vol. 9 (1933), pp. 31-34. I publish my proof in the following since it seems to be somewhat simpler.

² C. L. Siegel, *Acta Arithmetica*, vol. 1 (1935), pp. 83-86.

$$(1) \quad \theta = \sum_{\nu=0}^{n-1} f_{\nu} \chi_{\nu}.$$

LEMMA. *The character*

$$\theta - \chi_0 = \sum_{\nu=1}^{n-1} f_{\nu} \chi_{\nu}$$

of G can be expressed as a linear combination $\sum c_{\rho} \omega_{\rho}^*$ of characters ω_{ρ}^* of G which are induced by irreducible characters ω_{ρ} different from the 1-character of cyclic subgroups, such that the coefficients c_{ρ} are positive rational numbers with the denominator g .

Proof. Let A be a cyclic subgroup of order $a > 1$ of G . Let $\psi_0(\alpha) = 1$, $\psi_1(\alpha), \dots, \psi_{a-1}(\alpha)$ be the characters of A where α denotes a variable element of A . Then $\chi_{\nu}(\alpha)$ can be written as a linear combination

$$(2) \quad \chi_{\nu}(\alpha) = \sum_{\mu=0}^{a-1} h_{\nu\mu} \psi_{\mu}(\alpha)$$

with coefficients $h_{\nu\mu}$ which are non-negative rational integers. The orthogonality relations for group characters show that here

$$(3) \quad h_{\nu\mu} = (1/a) \sum_{\alpha} \overline{\chi_{\nu}(\alpha)} \psi_{\mu}(\alpha),$$

α ranging over all elements of A .

The character ψ_{μ} of A induces a character ψ_{μ}^* of G . According to a theorem of Frobenius,³ we have

$$(4) \quad \psi_{\mu}^*(\sigma) = \sum_{\nu=0}^{n-1} h_{\nu\mu} \chi_{\nu}(\sigma) \quad (\sigma \text{ in } G)$$

where the coefficients $h_{\nu\mu}$ in (4) are the same numbers as in (2).

Thus

$$(5) \quad \psi_{\mu}^*(\sigma) = (1/a) \sum_{\nu=0}^{n-1} \chi_{\nu}(\sigma) \sum_{\alpha} \overline{\chi_{\nu}(\alpha)} \psi_{\mu}(\alpha).$$

Determine now numbers u_0, u_1, \dots, u_{a-1} such that

$$(6) \quad \sum_{\mu=0}^{a-1} u_{\mu} \psi_{\mu}(\alpha) = \begin{cases} a\phi(a), & \text{if } \alpha = 1.^4 \\ 0, & \text{if } \alpha \neq 1, \text{ but does not generate } A. \\ -a, & \text{if } \alpha \text{ generates } A. \end{cases}$$

The orthogonality relations for the group characters of A make it possible to write down the u_{μ} explicitly. We have

³ G. Frobenius, *Sitzungsberichte der Berliner Akademie* (1898), pp. 501-515. Compare also A. Speiser, *Theorie der Gruppen von endlicher Ordnung*, 3rd ed. (1937), Berlin, § 64.

⁴ Here, $\phi(a)$ denotes the Euler function.

$$u_\mu = (1/a) (\phi(a) \overline{a\psi_\mu(1)} + \sum_{\beta} (-a) \overline{\psi_\mu(\beta)})$$

where β ranges over the primitive elements of A . Hence

$$(7) \quad u_\mu = \phi(a) - \sum_{\beta} \overline{\psi_\mu(\beta)}.$$

This shows that u_μ is a rational number and hence a rational integer. Further, since $|\psi_\mu(\beta)| = 1$ and since $\psi_\mu(\beta) \neq 1$ except for the 1-character ψ_0 , it follows from (7) that

$$(8) \quad u_0 = 0, u_\mu > 0 \text{ for } \mu > 0.$$

Form now the expression

$$(9) \quad S(A) = \sum_{\mu} u_{\mu} \psi_{\mu}^*(\sigma).$$

Then, by (5)

$$S(A) = (1/a) \sum_{\nu} \chi_{\nu}(\sigma) \sum_{\alpha} \overline{\chi_{\nu}(\alpha)} \sum_{\mu} u_{\mu} \psi_{\mu}(\alpha)$$

and (6) yields

$$S(A) = \sum_{\nu} \chi_{\nu}(\sigma) (\overline{\chi_{\nu}(1)} \phi(a) - \sum_{\beta} \overline{\chi_{\nu}(\beta)})$$

where β again ranges over the primitive elements of A . This shows that the function $S(A)$ of a variable element σ of G is a linear combination of the characters χ_{ν} . The value $\overline{\chi_{\nu}(1)}$ is the degree f_{ν} of χ_{ν} . In particular, the 1-character χ_0 appears in $S(A)$ with the coefficient

$$\phi(a) - \sum_{\beta} 1 = \phi(a) - \phi(a) = 0.$$

It is sufficient to let ν range over the values $1, 2, \dots, n-1$:

$$(10) \quad S(A) = \sum_{\nu=1}^{n-1} \chi_{\nu}(\sigma) (f_{\nu} \phi(a) - \sum_{\beta} \overline{\chi_{\nu}(\beta)}).$$

Add now the formulas (10) for *all* cyclic subgroups A of an order > 1 . Then β will range over all elements $\neq 1$ of G , and since χ_{ν} is not the 1-character χ_0 , the orthogonality relations for group characters yield

$$-\sum_A \sum_{\beta} \overline{\chi_{\nu}(\beta)} = \overline{\chi_{\nu}(1)} = f_{\nu}.$$

Further

$$\sum_A \phi(a) = g - 1.$$

Hence (10) gives

$$\sum S(A) = \sum_{\nu=1}^{n-1} \chi_{\nu}(\sigma) (f_{\nu}(g-1) + f_{\nu}) = g \sum_{\nu=1}^{n-1} f_{\nu} \chi_{\nu}(\sigma).$$

Thus

$$(11) \quad (1/g) \sum S(A) = \sum_{v=1}^{n-1} f_v \chi_v(\sigma) = \theta - \chi_0.$$

It is shown by (9) that the left side of (11) is a linear combination of characters ψ_μ^* induced by irreducible characters ψ_μ of cyclic subgroups. From (8), it follows that no character ψ_0^* induced by a 1-character appears, and that the other ψ_μ^* appear with positive coefficients u_μ/g . These coefficients are rational numbers with the denominator g . This establishes the lemma.

3. We now state

THEOREM 1. *Let k be an algebraic number field and let K be a normal field over k , both fields being of finite degrees. Then the quotient of the zeta-functions*

$$\zeta(s, K)/\zeta(s, k)$$

is an integral function of the complex variable s .

Proof. The zeta-functions can be expressed as L -series in the sense of Artin ⁵

$$\zeta(s, K) = L(s; \theta, K/k); \zeta(s, k) = L(s; \chi_0, K/k)$$

where θ now denotes the character of the regular representation of the Galois group G of K with regard to k and where χ_0 is the 1-character of G . Hence

$$\zeta(s, K)/\zeta(s, k) = L(s; \theta - \chi_0, K/k)$$

and Lemma 1 yields

$$(12) \quad \zeta(s, K)/\zeta(s, k) = \Pi L(s; \omega_\rho^*, K/k)^{e_\rho}.$$

According to a fundamental result of Artin,⁶ the L -series $L(s; \omega_\rho^*, K/k)$ is identical with an abelian L -series $L(s, \omega_\rho) = L(s; \omega_\rho, U/V)$ where U is a cyclic field over the field V ; $k \subseteq V \subset U \subseteq K$. Here, ω_ρ is a linear character, not the 1-character, and consequently ⁷ $L(s; \omega_\rho^*, K/k)$ is an integral function. Now (12) together with Lemma 1 shows that the g -th power of $\zeta(s, K)/\zeta(s, k)$ is an integral function. Since $\zeta(s, K)/\zeta(s, k)$ is meromorphic, it must be integral. This proves Theorem 1. At the same time we obtain:

⁵ See § 4 of Artin's paper in the *Abhandlungen aus dem Mathematischen Seminar Hamburg*, vol. 3.

⁶ See § 5 of Artin's paper referred to in ⁵. The general law of reciprocity was proved by Artin in *Abhandlungen aus dem Mathematischen Seminar Hamburg*, vol. 5 (1927), pp. 353-363.

⁷ This had first been shown by E. Hecke. See E. Landau, *Mathematische Zeitschrift*, vol. 2 (1918), pp. 52-154, Theorem LXIII.

COROLLARY. *If the normal field K has the degree g over k , then the g -th power of $\zeta(s, K)/\zeta(s, k)$ can be written as a product of abelian L -series which belong to linear characters different from the 1-character.*

It has been surmised that Theorem 1 remains valid, if K is an arbitrary (not normal) extension field of k of finite degree. This would follow from the still stronger conjecture of Artin that the L -functions belonging to irreducible characters different from the 1-character are integral functions. Our present method does not offer a possibility to attack these problems.⁸

4. We now wish to prove the following theorem formulated by Siegel.

THEOREM 2. *Let d denote the discriminant of the algebraic number field k of degree n over the field P of rational numbers. Let h be the number of classes of ideals in k and let R be the regulator. Then, for the fields k of fixed degree n*

$$\log(hR) \sim \log \sqrt{|d|} \quad (\text{for } |d| \rightarrow \infty).$$

Proof. (a) We state without proof a number of known results which will be used in what follows.

LEMMA 1 (Siegel⁹). *Let W be an algebraic number field of degree q over the field of rational numbers with the discriminant T . Denote by $r(W)$ the residue of the zeta-function $\zeta(s, W)$ for $s = 1$. If $\zeta(s, W) \leq 0$ for a real value of s with $0 < s < 1$, then*

$$(13) \quad r(W) > s(1-s)2^{-q}e^{-2q\pi} |T|^{(s-1)/2}.$$

Siegel's Lemma 4 must be replaced by

LEMMA 2 (Landau¹⁰). *Let W , q , T have the same significance as in Lemma 1. Let χ be a linear character, not the 1-character, modulo the ideal \mathfrak{f} in W . The corresponding (abelian) L -series $L(s, \chi)$ satisfies for $s = 1$ the inequality*

⁸ In the case of the icosahedral group where the characters of G are known explicitly, Artin could show that the zeta-function of the ground field divides the zeta-functions of some of the fields Ω lying between k and K , but not all these Ω can be treated in this manner. Actually, it is not known to-day whether $\zeta(s, k)$ divides $\zeta(s, \Omega)$ in these other cases.

⁹ Siegel, *loc. cit.*, Lemma 2. The proof of the lemma, as is that of the following two lemmas, is based on Hecke's integrals for the zeta-functions and the abelian L -series.

¹⁰ E. Landau, *Mathematische Zeitschrift*, vol. 4 (1919), pp. 152-162; theorem 5. While the corresponding lemma in Siegel's case can be proved quite directly in an elementary manner, it seems impossible at the present time to prove Lemma 3 avoiding Hecke's theory of analytic continuation of the abelian L -series.

$$(14) \quad |L(1, \chi)| \leq \lambda_q \log^q |T \cdot N\mathfrak{f}|$$

where λ_q depends on q alone and where $N\mathfrak{f}$ denotes the norm of \mathfrak{f} .

LEMMA 3 (Landau¹¹). *Let k be an algebraic number field of degree n over the field of rational numbers. If h denotes the number of classes of ideals in k , if d is the discriminant and R the regulator of k , then*

$$(15) \quad hR \leq \lambda'_n \sqrt{|d|} \log^{n-1} |d|$$

where λ'_n is a constant depending on n alone.

(b) We wish to show that for all algebraic number fields k of degree n the relation

$$(16) \quad \log r(k) = o(\log |d|)$$

holds where $r(k)$ is the residue of $\zeta(s, k)$ for $s=1$ and where d is the discriminant of k . As is well known,

$$(17) \quad r(k) = chR/\sqrt{|d|}$$

where c lies between positive bounds depending on n alone. Hence (15) shows that

$$(18) \quad \overline{\lim} \log r(k)/\log |d| \leq 0.$$

We now consider the case where k is a normal field of degree n over the field of rational numbers and apply Siegel's method. If (16) were false for normal fields k , there would exist a positive number $\epsilon < 1$ such that for infinitely many normal fields k of degree n the following inequality holds

$$(19) \quad r(k) < |d|^{-\epsilon}.$$

Set

$$(20) \quad \eta = 2\epsilon/n$$

and choose k such that (19) holds and that $|d|$ satisfies the inequality

$$(21) \quad |d|^{-\epsilon} < (1 - \eta)\eta 2^{-n} e^{-2n\pi} |d|^{-\eta/2}.$$

By Lemma 1 and (19) and (21), $\zeta(1 - \eta, k) > 0$. Since $\zeta(s, k) \rightarrow -\infty$ for real s approaching 1 from the left, it follows that there exists a real σ with

¹¹ E. Landau, *Göttingen Nachrichten* (1918), pp. 478-488, proof of Lemma 1.

$$(22) \quad 1 - \eta < \sigma < 1$$

such that

$$(23) \quad \xi(\sigma, k) = 0.$$

From now on, the field k will be considered as fixed. Let K be a second field, normal of degree n over the field of rational numbers, and let $\Omega = Kk$ be the compositum of K and k . Then the degree m of Ω divides n^2 . If K has the discriminant D , the discriminant Δ of Ω will divide $d^n D^n$.

It follows from the Corollary in §3 that we may set

$$(24) \quad \xi(s, \Omega) = \xi(s, k) M_0(s),$$

$$(25) \quad \xi(s, \Omega) = \xi(s, K) M(s).$$

Here, the n^2 -th powers of $M_0(s)$ and $M(s)$ are products of L -series $L(s; \chi, U/V)$ where U is a cyclic extension field of the field V , $V \subset U \subseteq \Omega$, and where χ is a linear character, different from the 1-character. The number of factors lies below a fixed bound depending on n . Then $L(s; \chi, U/V)$ is identical with an L -series $L(s, \chi)$ of the field V where χ now is a character modulo a suitable ideal \mathfrak{f} , the conductor, and χ is not the 1-character. It follows from the conductor-discriminant formula of class field theory¹² that $N\mathfrak{f}$ divides the discriminant of U which itself divides the discriminant Δ of Ω and hence $d^n D^n$. The discriminant of V also will divide $d^n D^n$. Take $|D| \geq |d|$. Lemma 2 implies that

$$|L(1, \chi, U/V)| \leq c_1 \log^{c_2} |D|$$

where c_1, c_2 (and later c_3, c_4, \dots) are positive constants depending on n alone. Then

$$(26) \quad |M(1)| \leq c_3 \log^{c_4} |D|.$$

It follows from (23) and (24) that $\xi(\sigma, \Omega) = 0$. Hence, by Lemma 1,

$$r(\Omega) > \sigma(1 - \sigma) 2^{-m} e^{-2m\pi} |\Delta|^{(\sigma-1)/2}$$

where m is the degree of Ω . Since $|\Delta| \leq |d|^n |D|^n$, this gives

$$(27) \quad r(\Omega) \geq a |D|^{-n(1-\sigma)/2},$$

a denoting a positive constant which depends on n, d, σ , but not on K . It follows from (25) that

¹² See, for instance, H. Hasse, *Klassenkörpertheorie*, mimeographed notes, Marburg 1932/33, p. 118 and §23. Actually, we use only a weaker result. This, as is well known, can be obtained analytically by comparing Hecke's functional equations for $\xi(s, U)$ and for the L -series belonging to the cyclic field U over V .

$$r(\Omega) = r(K)M(1).$$

Combining this with (26) and (27), we obtain

$$r(K) = r(\Omega)/M(1) \geq ac_3^{-1} |D|^{-n(1-\sigma)/2} \log^{-c_4} |D|.$$

Since $1 - \sigma < \eta = 2\epsilon/n$, it follows that

$$r(K) \geq |D|^{-\epsilon}$$

for all normal fields K of degree n with sufficiently large $|D|$. This contradicts the assumption that the inequality (19) holds for infinitely many normal fields of degree n , and this shows that (16) holds for normal fields k .

(c) Consider now an arbitrary number field k of degree n with the discriminant d . Again, we wish to establish the relation (16). Let Ω denote now the Galois field belonging to k . Then Ω is normal over the field of rational numbers. Its degree lies below a fixed constant c_5 depending on n only. If Δ is the discriminant of Ω , then $|\Delta| \leq |d|^\nu$ where ν depends only on n . On account of the Corollary in §3, we may set

$$(28) \quad \zeta(s, \Omega) = \zeta(s, k)M^*(s)$$

where $M^*(s)$ has properties analogous to those of $M(s)$ in (25). Applying the argument leading to (26) here, we obtain now

$$|M^*(1)| \leq c_6 \log^{c_7} |d|.$$

Now (28) yields

$$(29) \quad r(k) = r(\Omega)/M^*(1) \geq r(\Omega)c_6^{-1} \log^{-c_7} |d|.$$

Since (16) has been established for normal fields of a fixed degree, we have $r(\Omega) \geq |\Delta|^{-\delta}$ for every fixed $\delta > 0$ provided that $|\Delta|$ is sufficiently large. But $|d| \leq |\Delta| \leq |d|^\nu$ and hence

$$(30) \quad r(\Omega) \geq |d|^{-\delta\nu}$$

provided that $|d|$ is sufficiently large. Combination of (29) and (30) gives

$$\liminf \log r(k) / \log |d| \geq 0.$$

This, together with (18), shows that (16) holds for all fields k of degree n .

Now (17) implies that $\log(hR) \sim \log \sqrt{|d|}$ as was stated in Theorem 2.

ASYMPTOTIC INTEGRATIONS OF THE ADIABATIC OSCILLATOR.*

By AUREL WINTNER.

1. The following considerations deal with the behavior, as $t \rightarrow \infty$, of the linear, non-conservative system defined by the Hamiltonian function

$$(1) \quad H(x, x'; t) = \frac{1}{2}\{x'^2 + (\omega^2 + \phi(t))x^2\}$$

of a single degree of freedom, that is, by the Lagrangian equation

$$(2) \quad x'' + (\omega^2 + \phi(t))x = 0,$$

where ω denotes a positive constant ("undisturbed frequency"), and the coefficient function $\phi(t)$, a given perturbation of the squared frequency, is "small" or "slow" as $t \rightarrow \infty$. It will always be assumed that the coefficient function, $\phi(t)$, of (2) is defined for large positive t , and that $\phi(t)$ is continuous and real-valued (these assumptions will not always be repeated in the wording of the theorems).

Actually, it will be convenient (and, in the proof of (i), almost necessary) to consider complex-valued coefficient functions also. Such coefficient functions will not be denoted by $\phi(t)$ but by other symbols. It will be understood that these coefficient functions, too, are defined and continuous for large positive t .

The "smallness" of $\phi(t)$, as $t \rightarrow \infty$, can be specified in various senses. The problem is precisely which of these specifications leads to an assigned restriction of the asymptotic behavior of the general solution of (2). The most natural of these specifications seems to be

$$(3) \quad \phi(t) \rightarrow 0 \text{ as } t \rightarrow \infty,$$

but this turns out to be quite useless, since it proves to be compatible with almost any asymptotic behavior of the solutions $x(t)$ (except, of course, with a behavior excluded by Sturm's comparison theorem on nodes); cf. [4]. Less primitive specifications result if $\phi(t)$ is required to be of class $(L) = (L^1)$

* Received November 7, 1946.

or of class (L^2) . Neither of these specifications implies the other and both of them together do not imply (3). It is understood that, since $\phi(t)$ is continuous on the half-line, it is said to be of class (L^p) if

$$(3_p) \quad \int_0^\infty |\phi(t)|^p dt < \infty \quad (p > 0).$$

Still another condition is

$$(3_0) \quad \int_0^\infty |d\phi(t)| < \infty$$

(to be required for a sufficiently large lower limit of integration). Since (3_0) , in contrast to (3_p) , does not imply that $\phi(t)$ is "small" for large t , the restriction (3) will always be assumed, if (3_p) , where $p > 0$, is replaced by (3_0) . In this regard, it is sufficient to observe that (3_0) implies the existence of a finite limit $\phi(\infty)$, which can have an arbitrary value, and that (3) means the vanishing of this limit.

It is convenient to think of (3_0) and (3) together as representing a limiting case of (3_p) . This limiting case of (L^p) will be denoted by (L^0) . In other words, $\phi(t)$ will be called of class (L^0) if it is of bounded variation (on some half-line) and tends to 0 as $t \rightarrow \infty$. For instance, a monotone $\phi(t)$ is of class (L^0) if and only if it satisfies (3). Needless to say, neither of the classes (L^0) , (L^p) contains the other (whether $p > 0$ is fixed or variable). All that is true is that $\phi(t)$ must be of class (L^0) if it satisfies (3_0) and is of class (L^p) (for some $p > 0$).

2. The results to be obtained are *explicit rules for the asymptotic integration of (2)*, the description of which can best be centered about the following theorem:

(i) *If $\phi(t)$ is of class (L^0) and of class (L^2) , then there belongs to every solution $x(t) \not\equiv 0$ of (2) a unique pair of positive integration constants, say a and $\alpha (\leq 2\pi)$, such that*

$$(4) \quad x(t) = a \cos(\alpha + \omega t + \tfrac{1}{2} \int_0^t \phi(s) ds / \omega) + \epsilon(t),$$

where the remainder term, $\epsilon(t) = \epsilon_{aa}(t)$, is $o(1)$, as is its derivative; that is,

$$(5) \quad \epsilon(t) \rightarrow 0 \text{ and } \epsilon'(t) \rightarrow 0 \text{ as } t \rightarrow \infty.$$

If the zeros of the cosine in (4) are disregarded, the first of the relations

(5) states that the first term on the right of (4) is asymptotic to the function on the left of (4), and the second of the relations (5) means that the formal differentiation of this asymptotic formula for $x(t)$ is legitimate.

Since (4) contains two integration constants, a and α , the lower limit of the quadrature occurring in (4) can be thought to be fixed. On the other hand, the fact that this lower limit is arbitrary cannot dispose of the necessity of an α in (4), since the function of t representing the third term beneath the cos can have a range the length of which is less than 2π . But this precaution is surely superfluous unless the coefficient function $\phi(t)$ of (2) is such that

$$(6) \quad \int_0^t \phi(s) ds$$

becomes $O(1)$ as $t \rightarrow \infty$.

For the still more restrictive case, in which (6) tends to a finite limit, (i) clearly contains the following corollary:

(i bis) *If $\phi(t)$ is of finite total variation and such as to make both improper integrals*

$$(7) \quad \int_0^\infty \phi^2(t) dt, \quad \int_0^\infty \phi(t) dt \equiv \lim_{T \rightarrow \infty} \int_0^T \phi(t) dt$$

convergent (whereas $\int_0^\infty |\phi(t)| dt = \infty$ is allowed), then there belongs to every solution, $x(t)$, of (2) a unique pair of integration constants, c_1 and c_2 , such that

$$(8) \quad x(t) = c_1 \cos \omega t + c_2 \sin \omega t + \epsilon(t),$$

where the remainder term and its derivative satisfy (5).

Actually, the (L^2) -assumption can now be omitted, since it is implied by the remaining assumptions of (i bis). In order to see this, it is sufficient to apply to the first of the integrals (7) an integration by parts (based on the factorization $\phi^2(t) dt = \phi(t) dF(t)$, where F denotes the indefinite integral of ϕ).

What is essential in (i bis) is the inclusion of the case mentioned parenthetically after (7). In fact, if $\phi(t)$ is of class (L^1) , then (8) and (5) hold for all solutions of (2) even if neither of the remaining assumptions of (i bis), requiring that $\phi(t)$ be of class (L^0) and of class (L^2) , is satisfied. Cf. the case $f(t) \equiv \omega^2$ of (iii) below.

A simple instance, to which (i) is applicable but (i bis) is not, results on choosing $\phi(t) = t^{-\lambda}$, where $\frac{1}{2} < \lambda \leq 1$. In fact, $\phi(t)$ then is of class (L^0) and of class (L^2) , but (6) is not $O(1)$. According to (4), the general solution of

$$(9) \quad x'' + (1 + t^{-\lambda})x = 0$$

is of the form

$$(10) \quad x(t) = c_1 \cos(t + \tfrac{1}{2} \log t) + c_2 \sin(t + \tfrac{1}{2} \log t) + \epsilon(t)$$

if $\lambda = 1$, and of the form

$$(11) \quad x(t) = c_1 \cos(t + t^\mu/\mu) + c_2 \sin(t + t^\mu/\mu) + \epsilon(t), \text{ where } \mu = 1 - \lambda,$$

if $\frac{1}{2} < \lambda \neq 1$; hence, of the form (8) if and only if $\mu < 0$, i. e., $\lambda > 1$.

3. Instead of the limiting case of the range $\frac{1}{2} < \lambda \leq 1$, covered above, consider the following modification of the case $\lambda = \frac{1}{2}$ of (9):

$$(12) \quad x'' + (1 + 2t^{\frac{1}{2}} + t^{-1})x = 0;$$

so that, in (2),

$$(13) \quad \phi(t) = 2t^{\frac{1}{2}} + t^{-1}$$

and $\omega = 1$. Since (13) means that

$$(14) \quad \tfrac{1}{2} \int^t \phi(s) ds = 2t^{\frac{3}{2}} + \tfrac{1}{2} \log t,$$

the assumption that (i) is applicable would mean that the general solution of (12) is a superposition of the real and imaginary parts of

$$(15) \quad \exp i(t + 2t^{\frac{3}{2}} + \tfrac{1}{2} \log t) + \epsilon(t).$$

In fact, (14) shows that this is equivalent to the assertion, (4), of (i), the value of ω being 1. But it will be shown that the general solution of (12) is a superposition of the real and imaginary parts of

$$(16) \quad \exp i(t + 2t^{\frac{3}{2}}) + \epsilon(t),$$

rather than of (15). It is understood that $\epsilon(t)$ in (15) and (16) denote two functions which (along with their derivative) satisfy (5). Hence, no superposition ($\neq 0$) resulting from (16) can result from (15). Consequently, the assumption that (i) is applicable to (12) is false.

Since the function (13) is of class (L^0) but fails to be (though just barely) of class (L^2) , it follows that the second of the assumptions of (i) cannot be omitted:

(i*) *The assertions of (i) are not in general true for a coefficient function, $\phi(t)$, which is of class (L^0) (but not, of course, of class (L^2) as well).*

In order to prove this, it will be convenient to start with the principal term,

$$(16^*) \quad x(t) = \exp i(t + 2t^{\frac{1}{2}}),$$

of (16) and to deduce for (16^*) a differential equation, say (12^*) , which differs from (12) only in unessential terms (to be specified in a moment), rather than to start with (12) and then deduce (16), which differs from (16^*) only in unessential terms (as specified by (5)).

First, if $x = x(t)$ is defined by (16^*) , then $x' = i(1 + t^{\frac{1}{2}})x$, hence

$$x'' = i^2(1 + t^{\frac{1}{2}})^2x - \frac{1}{2}(t^{\frac{1}{2}})^3ix.$$

This can be written in the form

$$(12^*) \quad x'' + (1 + 2t^{\frac{1}{2}} + t^{-1} + \frac{1}{2}(t^{\frac{1}{2}})^3i)x = 0,$$

which is (2) with $\omega = 1$ and with ϕ^* instead of $\phi(t)$, where

$$(13^*) \quad \phi^*(t) = \phi(t) + \frac{1}{2}(t^{\frac{1}{2}})^3i,$$

if $\phi(t)$ denotes the coefficient function, (13), of (12).

Accordingly, the deviation of the coefficient functions of (12) and (12^*) , though imaginary, is $O(t^{-\frac{1}{2}})^3$, and therefore of class (L^1) . But (iii) below will show that a deviation of class (L^1) in the coefficient functions, whether real-valued or not, does not have any effect on the asymptotic form of the general solution; in the sense that the whole disturbance will consist of an additive correction term which (along with its derivative) satisfies (5). Since, in terms of the superposition of real and imaginary parts, (16^*) belongs to (12^*) , it follows that what belongs to (12) is (16), rather than (15).

This proves (i*). In order to ascertain that both assumptions of (i) are essential, the following dual of (i*) remains to be verified:

(i**) *The assertions of (i) are not in general true for a coefficient function, $\phi(t)$, which tends, as $t \rightarrow \infty$, to 0 and is of class (L^2) (but not, of course, of class (L^0) as well).*

In fact, Perron [2] has exhibited a (real-valued) coefficient function $\phi(t)$ and a solution $x = x(t)$ of the differential equation (2), belonging to this $\phi(t)$ which have the following properties: $\phi(t) = O(t^{-1})$ and $x(t) \neq O(1)$, as $t \rightarrow \infty$ (for a simple deduction and for more general constructions, cf. [4], pp. 386-387 and pp. 394-396). This example proves (i**), since, on the one hand, $\phi(t) = O(t^{-1})$ implies that $\phi(t)$ is of class (L^2) and satisfies (3) and, on the other hand, $x(t) \neq O(1)$ contradicts the assertion, (4), of (i).

4. What is known under the (L^0) -assumption alone can be summarized as follows:

(ii₀) If $\phi(t)$ is of class (L^0) , then, although (4) and (5) need not hold,

$$(17) \quad x(t) = O(1) \text{ and } x'(t) = O(1) \text{ as } t \rightarrow \infty$$

must hold, for every solution, $x = x(t)$, of (2).

The first assertion of (ii₀) is precisely (i*). The second claims that (3) and (3₀) imply (17) for every solution of (2). This has been proved by A. Kneser [1], pp. 73-80. In fact, only his wording, but not his proof, of (17) assumes more than the pair of conditions (3), (3₀). But the proof is made unnecessarily involved by Kneser's use of Sturm's oscillation theorem. If the latter is avoided, more than (17) can be proved:

(ii) If $\phi(t)$ is of class (L^0) and if

$$(18) \quad h(t) = H(x(t), x'(t); t)$$

denotes the energy, (1), along any solution, $x = x(t)$, of (1), then

$$(19) \quad \int_0^\infty |dh(t)| < \infty.$$

COROLLARY. The energy, (18), of every solution $x = x(t)$ tends, as $t \rightarrow \infty$, to a finite limit,

$$(20) \quad h(\infty).$$

The latter is positive except when

$$(20 \text{ bis}) \quad x(t) \equiv 0, \text{ hence } h(t) \equiv 0.$$

In order to obtain (17), not even the existence of a finite limit, (20), is needed, since $h(t) = O(1)$ is sufficient to this end. In fact, $h(t) = O(1)$ means, by (18) and (1), that

$$x'^2(t) + (\omega^2 + \phi(t))x^2(t) = O(1),$$

which, in view of (3), is equivalent to (17).

Remark. Let a_n denote the n -th amplitude, and b_n the n -th co-amplitude, of a (real-valued) solution $x = x(t) \not\equiv 0$ (as to terminology, cf. [4], pp. 389-391). Then it is easily seen from (21⁰) below, from the notations (18) and (1), and from (3), that the assertion, (19), of (ii), implies

$$(21) \quad \sum_{n=1}^{\infty} |a_{n+1} - a_n| < \infty \quad \text{and} \quad \sum_{n=1}^{\infty} |b_{n+1} - b_n| < \infty.$$

Kneser's result, according to which there exist finite positive limits

$$(21_0) \quad \lim_{n \rightarrow \infty} a_n \quad \text{and} \quad \lim_{n \rightarrow \infty} b_n$$

if $x(t) \not\equiv 0$ ([1], p. 80) is, of course, a corollary of (21). Finally,

$$(21^0) \quad \lim_{n \rightarrow \infty} a_{n+1}/a_n = 1 \quad \text{and} \quad \lim_{n \rightarrow \infty} b_n/a_n = \omega$$

hold even if $\phi(t)$, instead of satisfying both (3) and (3₀), satisfies (3) only (cf. [4], p. 391).

5. The proof of (ii) can be based on the fact that, in virtue of (18) and (1),

$$(22) \quad dh(t) = \frac{1}{2}x^2(t)d\phi(t)$$

is an identity in t along any solution, $x = x(t)$, of (2). This fact is a straight-forward generalization of the energy relation, $h(t) = \text{const.}$, of the conservative case, $\phi(t) \equiv 0$, and is verified in the same way, that is, by direct substitution and partial integration. Caution is necessary only if $\phi(t)$ is just continuous. In the sense of Stieltjes integrations, (22) is true in this case also, it being understood that (22) must then be interpreted as follows:

$$(22 \text{ bis}) \quad \int_u^v g(t) dh(t) = \frac{1}{2} \int_u^v g(t) x^2(t) d\phi(t)$$

holds for arbitrary u, v , and for every $g(t)$ of bounded variation.

It is clear from (18), (1) and (3) that, from a certain t onward,

$$(23) \quad x^2(t) + x'^2(t) < h(t),$$

if the unit of length on the t -axis is so chosen that $4\omega^2$ becomes 1. Let the

trivial solution, (20 bis), be excluded. Then $x(t)$ and $x'(t)$ cannot vanish simultaneously and so, by (23),

$$(23 \text{ bis}) \quad h(t) > 0.$$

But (22) and (23) imply that, corresponding to the interpretation, (22 bis), of (22),

$$|dh(t)| \leq h(t) |d\phi(t)|; \text{ hence, } |d \log h(t)| \leq |d\phi(t)|,$$

by (23 bis). It follows, therefore, from (3₀) that

$$\int^{\infty} |d \log h(t)| < \infty.$$

This implies that $\log h(t)$ tends to a finite limit as $t \rightarrow \infty$. In particular, $\log h(t) = O(1)$, and so both $h(t)$ and $1/h(t)$ are $O(1)$. Hence, the result of the last formula line is equivalent to the assertion, (19), of (ii). Finally, the vanishing of (20) is precluded by $1/h(t) = O(1)$; so that $h(\infty) > 0$, by (23 bis).

5 bis. More interesting than the set of relations (21), (21₀), (21°) is the following fact (which, as will be shown elsewhere, is needed in an application to the theory of spectra):

(ii bis) *If $\phi(t)$ is of class (L^0), and if t_1, t_2, \dots denotes the sequence of consecutive zeros of an arbitrary solution $x(t) \not\equiv 0$ of (2), then the mean square of $x(t)$ over a half-wave, that is, the average*

$$\int_{t_n}^{t_{n+1}} x^2(t) dt / (t_{n+1} - t_n),$$

(as well as $t_{n+1} - t_n$ itself), tends to a finite, positive limit as $n \rightarrow \infty$.

In order to simplify the notations, let the units of length on the x - and t -axes be so chosen that the integration constant (20) (which is positive, since (20 bis) is excluded) and the constant ω occurring in (2) become 1. Since, by (3),

$$\omega^2 + \phi(t) = \omega^2 + o(1) = 1 + o(1) \text{ as } t \rightarrow \infty,$$

"the nodes of $x(t)$ are majorized and minorized," in Sturm's sense, by the nodes of the non-trivial solutions of the two linear oscillators

$$y'' + (1 \pm \epsilon)y = 0,$$

where ϵ is an arbitrarily fixed positive constant and t is sufficiently large (viz., $T < t < \infty$, where T depends only on ϵ). Since the n -th half-waves of the solutions $y(t) \not\equiv 0$ of these linear oscillators are of length $\pi/(1 \pm \epsilon)^{\frac{1}{2}}$ for every n , it follows that every solution $x(t) \not\equiv 0$ of (2) has a sequence of zeros, and that, if t_1, t_2, \dots , where $t_n < t_{n+1}$, denotes the sequence of all these zeros (when $x(t)$ is fixed), then

$$t_{n+1} = t_n + \pi + o(1) \text{ as } n \rightarrow \infty.$$

Since $x(t)$ vanishes at $t = t_n$ and at $t = t_{n+1}$, a partial integration gives

$$\int_{t_n}^{t_{n+1}} -x''(t)x(t)dt = \int_{t_n}^{t_{n+1}} x'^2(t)dt.$$

But (2), where $\omega = 1$, shows that

$$-x''(t)x(t) = x^2(t)[1 + \phi(t)] = x^2(t) + o(1) \text{ as } t \rightarrow \infty,$$

by (3) and by the first of the estimates (17). Hence, the integral on the left becomes

$$\int_{t_n}^{t_{n+1}} \{x^2(t)dt + o(1)\}dt$$

as $n \rightarrow \infty$. On the other hand, since the constant (20) has been chosen to be 1, it is seen from (18) and (1), where $\omega = 1$, that, as $t \rightarrow \infty$,

$$x'^2(t) + x^2(t) = 2 + o(1),$$

again by (3) and by the first of the relations (17). Hence, the integral on the right is

$$\int_{t_n}^{t_{n+1}} \{2 - x^2(t) + o(1)\}dt.$$

Consequently,

$$\int_{t_n}^{t_{n+1}} x^2(t)dt + \int_{t_n}^{t_{n+1}} o(1)dt = \int_{t_n}^{t_{n+1}} 2dt - \int_{t_n}^{t_{n+1}} x^2(t)dt.$$

In view of $t_{n+1} - t_n \rightarrow \pi$, where $n \rightarrow \infty$, this means that

$$\int_{t_n}^{t_{n+1}} x^2(t) dt = \pi + \int_{t_n}^{t_{n+1}} o(1) dt + o(1) = \pi + o(1)$$

and completes, therefore, the proof of (ii bis).

6. In order to abbreviate the formulae, the following proof of the *asymptotic* assertions of (i) will be carried out under the unessential (since *local*) assumption that $\phi(t)$ has a continuous derivative, $\phi'(t)$. Without this simplification, the proof ought to work with Stieltjes differentials or, rather, with partial integrations which correspond to the interpretation, (22 bis), of (22) in the above proof of (ii).

Clearly, the assertion of (i) is equivalent to the statement that every solution of

$$(24) \quad x'' + (1 + \phi(t))x = 0$$

is a superposition of the real and imaginary parts of a (particular) complex-valued solution of the form

$$(25) \quad x(t) = y(t) + \epsilon(t),$$

where $y(t)$ is an abbreviation for

$$(26) \quad y(t) = \exp i(t + \tfrac{1}{2} \int^t \phi(s) ds)$$

and $\epsilon(t)$ is some function which, together with its derivative, satisfies (5). Correspondingly, (i) will be proved along the lines of the above proof of (i*); that is, the function (26), which becomes a solution, (25), only after the addition of a correction term, will first be substituted into the differential operator on the left of (24), leading to an error in (24). The deviation from 0 of the value of the differential operator for $y(t)$ will then be estimated so as to lead to (5). This will be accomplished by an appropriate application of Lagrange's principle of "varied constants."

First, the derivative of the function (26) is

$$y' = i(1 + \tfrac{1}{2}\phi)y.$$

If this representation of y' is differentiated (and, by the assumption made before (24), it *can* be differentiated), it follows that

$$y'' = i^2(1 + \tfrac{1}{2}\phi)^2y + \tfrac{1}{2}i\phi'y.$$

Clearly, the last relation can be written in the form

$$(28) \quad y'' + (1 + \phi(t) + \psi(t))y = 0,$$

if ψ denotes the (complex-valued) function defined by

$$(29) \quad 2\psi(t) = \frac{1}{2}\phi^2(t) - i\phi'(t).$$

By the assumption made before (24), the function (29) is continuous. Furthermore, by the assumptions of (i), the function $\phi(t)$ belongs to both classes (L^0) , (L^2) ; that is, both (3_0) and (3_2) are satisfied. But (3_0) means that the second, and (3_2) that the first, term on the right of (29) is of class (L^1) . Accordingly, $\psi(t)$ is a (complex-valued) continuous function of class (L^1) .

Since (26) is a solution of (28), and since what (i) claims for (24) is (25) with (5), that is,

$$(30) \quad x(t) - y(t) \rightarrow 0 \text{ and } x'(t) - y'(t) \rightarrow 0 \text{ as } t \rightarrow \infty,$$

it is clear that (i) will be proved if it is shown that (30) represents a one-to-one correspondence between the solutions, $x(t)$, of (24) and the solutions, $y(t)$, of (28).

For the sake of short formulations, it will be convenient to use the following manner of speaking:

(iii⁰) *A continuous (possibly complex-valued) function, $f(t)$, defined for large positive t , will be called of class (*) if*

$$x(t) = O(1) \text{ and } x'(t) = O(1) \text{ as } t \rightarrow \infty$$

holds for every solution of the differential equation $x'' + f(t)x = 0$.

In terms of this definition, all that remains to be proved is contained in the following lemma:

(iii) *If $f(t)$ and $g(t)$ are continuous (possibly complex-valued) functions defined for large positive t in such a way that $f(t)$ is of class (*) and $g(t) - f(t)$ of class (L^1) , then (30) represents a one-to-one correspondence between the solutions of*

$$(31) \quad x'' + f(t)x = 0$$

and the solutions of

$$(32) \quad y'' + g(t)y = 0.$$

If (iii) is granted, the proof of (i) can be completed by identifying (24) with (31), and (28) with (32). In fact, $g(t) - f(t)$ then becomes the

function $\psi(t)$ which, as verified after (29), is of class (L^1) . Hence, it is sufficient to ascertain that also the other assumption of (iii), that requiring that $f(t)$ be of class $(*)$, is satisfied. But $f(t)$ now is the function $1 + \phi(t)$ which, in view of (iii⁰) and of the by-product, (17), of the Corollary of (ii), is of class $(*)$, since, by the assumptions of (i), the function $\phi(t)$ is of class (L^0) .

7. What remains to be ascertained, viz., the truth of (iii), is a particular case of a general theorem, appearing elsewhere, which concerns arbitrary systems of linear differential equations of first order. Due to the (formally) self-adjoint character of the (possibly complex) differential equations (32), (31), the proof of (iii) itself affords certain simplifications (and leads, in addition, to a refinement of (iii); cf. (iv) below). It will be given in a form which, instead of the classical procedure of successive approximations, depends only on an adaptation of the trivial estimates used in the proof of (ii).

Let $x = u(t)$ and $x = v(t)$ be two linearly independent solutions of (31). Since their Wronskian, $u(t)v'(t) - v(t)u'(t)$, is a non-vanishing constant (Abel), it can be assumed to be 1. Then

$$X^{-1}(t) = \begin{pmatrix} v'(t) & -v(t) \\ -u'(t) & u(t) \end{pmatrix}, \text{ if } X(t) = \begin{pmatrix} u(t) & v(t) \\ u'(t) & v'(t) \end{pmatrix}.$$

Hence, if

$$A(t) = \begin{pmatrix} 0 & 1 \\ -f(t) & 0 \end{pmatrix} \text{ and } B(t) = \begin{pmatrix} 0 & 1 \\ -g(t) & 0 \end{pmatrix},$$

then

$$X^{-1}(B - A)X = (f - g) \begin{pmatrix} -uv & -v^2 \\ u^2 & uv \end{pmatrix}.$$

But (31), (32) can be written as $\xi' = A(t)\xi$, $\eta' = B(t)\eta$, if ξ , η denote the binary vectors the first and second components of which are x , x' and y , y' , respectively. On the other hand, Lagrange's rule for the variation of constants (which can, of course, be verified by direct substitution of the matrices and vectors involved) states that, by virtue of the Wronskian transformation, $X(t)$, of the binary system $\xi' = A(t)\xi$, the binary system $\eta' = B(t)\eta$ is equivalent to

$$\xi' = C(t)\xi, \quad C = X^{-1}(B - A)X,$$

if $\zeta = \zeta(t)$ denotes the binary vector defined by $\zeta = X^{-1}(t)\eta$, where $\eta = \eta(t)$ is an *arbitrary* solution of $\eta' = B(t)\eta$.

Let $p = p(t)$ denote the first, and $q = q(t)$ the second, component of $\zeta = \zeta(t)$. Then the last two formula lines mean that

$$(33) \quad p' = (g - f)(uvp + v^2q), \quad q' = (f - g)(u^2p + uvq).$$

On the other hand, $X(t)$ having been defined to be the Wronskian matrix of $u = u(t)$ and $v = v(t)$, the definition, $\zeta = X^{-1}(t)\eta$, of ζ means that

$$(34) \quad y = up + vq, \quad y' = u'p + v'q.$$

Since $x = u(t)$ and $x = v(t)$, being solutions of (31), are assumed to be bounded *along with their derivatives*, $u'(t)$ and $v'(t)$, and since the determinant of the linear transformation, (34), of (p, q) into (y, y') is the Wronskian, which is 1 for every t , finally, since every solution, $x = x(t)$, of (31) is a unique superposition of the two particular solutions, $x = u(t)$ and $x = v(t)$, which have been selected for $X(t)$, it is clear that the assertion of (iii) is equivalent to the following statement: If $p = p(t)$, $q = q(t)$ is any solution of the binary linear system (33) (in which u and v are *given* functions of t , as are f and g), then $p(t)$, $q(t)$ tend, as $t \rightarrow \infty$, to finite limits,

$$(35) \quad p(\infty), q(\infty),$$

both of which vanish only for the trivial solution,

$$(35 \text{ bis}) \quad p(t) \equiv 0 \equiv q(t),$$

of (33).

8. In order to prove the existence of the finite limits (35) for every solution of (33), exclude the trivial solution (35 bis). Then, by the uniqueness theorem of linear systems of differential equations, $p(t)$ and $q(t)$ cannot vanish simultaneously. Hence, if $r(t)$ denotes the function

$$(36) \quad r(t) = (|p(t)|^2 + |q(t)|^2)^{\frac{1}{2}}$$

(which is positive, whereas $p(t)$ and $q(t)$ can be complex-valued), then the derivative $r'(t)$ exists ($\neq \pm \infty$), since, according to (33), both $p'(t)$ and $q'(t)$ exist.

On adding the relations which result if the first of the equations (33) is multiplied by p and the second by q , and then taking the analogous com-

binations belonging to complex conjugates, one readily sees from (36) that, since $2|u(t)v(t)| \leq |u(t)|^2 + |v(t)|^2$,

$$|r(t)r'(t)| \leq \text{Const.} |f(t) - g(t)| r^2(t) (|u(t)|^2 + |v(t)|^2)$$

holds for a certain absolute constant. In view of $r(t) > 0$, this inequality implies that, as $t \rightarrow \infty$,

$$(37) \quad |(\log r(t))'| = O(1) |f(t) - g(t)| (|u(t)|^2 + |v(t)|^2).$$

But (iii)^o shows that what is required of $f(t)$ in (iii) is that all four functions

$$(38) \quad u(t), v(t); u'(t), v'(t)$$

be $O(1)$, since $x = u(t)$ and $x = v(t)$ are two (linearly independent) solutions of (30). On the other hand (even if just the first two of the four functions (37 bis) are $O(1)$ as $t \rightarrow \infty$), it follows from (37) that

$$(39) \quad \log r(t) = O(1) \int^t |f(s) - g(s)| ds = O(1),$$

by the assumption made in (iii) for $f(t) - g(t)$. Since (39) implies that $r(t) = O(1)$, it now follows from (36) that both $p(t)$ and $q(t)$ are $O(1)$. Hence, it is seen from (33) that

$$(40) \quad p'(t) = O(1) |f(t) - g(t)|, \quad q'(t) = O(1) |f(t) - g(t)|,$$

since the functions (38) are $O(1)$ (actually, only the first two of the four functions (38) are needed).

Since the integral occurring in (39) is $O(1)$, it follows from (40) that both integrals

$$(41) \quad \int^{\infty} p'(t) dt, \quad \int^{\infty} q'(t) dt$$

are absolutely convergent. In particular, they are convergent. But their convergence proves the existence of the finite limits (35).

Finally, at least one of the two values (35) is distinct from 0. For, if the contrary is assumed, it follows from (36) that $r(t) \rightarrow 0$ as $t \rightarrow \infty$. Since this implies that $\log r(t) \neq O(1)$, it contradicts (39).

9. This proves (iii), hence (i) as well, and so, since (ii) has been proved earlier, all of the statements made above.

The method applied leads somewhat further than (iii). In order to

express the resulting refinement conveniently, use will be made of the following definition:

(iv⁰) Let a continuous function $f(t)$, defined for large positive t , be called of class $[*]$ if $x(t) = O(1)$ holds for every solution of $x'' + f(t)x = 0$.

Thus, in contrast to (iv⁰), it is now not required that $x'(t) = O(1)$. Correspondingly, the generalization alluded to before is as follows;

(iv) In (iii), the assumption that $f(t)$ be of class $(*)$ can be relaxed to the assumption that $f(t)$ be of class $[*]$, if (30) in the assertion of (iii) is relaxed to

$$(30 \text{ bis}) \quad x(t) - y(t) \rightarrow 0 \text{ as } t \rightarrow \infty.$$

In accordance with the remark made after (iv⁰), only the first of the two relations (30) is involved in the one-to-one correspondence claimed in (iv). Correspondingly, the truth of (iv) can be seen as follows:

In the preceding proof of (iii), the assumption requiring $f(t)$ to be of class $(*)$ has been used three times, the first time after (34), the second time before (39) and the last time after (40). However, the estimate $O(1)$ has been used only for the first two, rather than for all four, of the functions (38). But the assumption, (iv⁰), of (iv) means that the first two of the functions (38) are $O(1)$. Hence, the proof of (iii) makes it clear that, under the assumptions of (iv), both components, $p(t)$ and $q(t)$, of every solution of (33) tend to finite limits, (35), both of which vanish only in the case (35 bis). Hence, if $p(\infty) = c_1$ and $q(\infty) = c_2$, the first of the relations (34) gives

$$y(t) = u(t)(c_1 + o(1)) + v(t)(c_2 + o(1)).$$

Since $u(t) = O(1)$ and $v(t) = O(1)$, this can be written in the form

$$(42) \quad y(t) = x(t) + o(1), \text{ where } x(t) = c_1 u(t) + c_2 v(t).$$

In order to complete the proof of (iv), all that remains to be ascertained is that the *single-valued* correspondence established by (42) is a *schlicht* correspondence between the solutions of (31) and (32). In other words, it is sufficient to assure that, by virtue of (42), where $(u(t), v(t))$ denotes a fixed pair of two linearly independent solutions of (31), the same pair of integration constants, (c_1, c_2) , cannot belong to two distinct solutions, say $y = y^*(t)$ $y = y^{**}(t)$, of (32). But, if this were not true, $y = y^* - y^{**}$ would be a solution $y(t) \not\equiv 0$ of (32) satisfying (42) with $c_1 = 0 = c_2$. This contains, however, a contradiction. In fact, $c_1 = 0 = c_2$ means that $p(\infty) = 0$

$= q(\infty)$, hence $p(t) \equiv 0 \equiv q(t)$, and so $y(t) \equiv 0$, by the first of the relations (34). This contradiction completes the proof of (iv).

It remains to be seen that (iv) actually is more general than (iii):

(iv*) *The coefficient function, $f(t)$, of (31) can be of class (*) without being of class [*]. But this cannot happen if $f(t) = O(1)$.*

In particular, it cannot happen for (2) if $\phi(t)$ satisfies (3).

First, two differentiations show that

$$x(t) = \exp(-t + i \exp 2t)$$

is a solution of (31) when

$$f(t) = -1 + 4 \exp 4t.$$

Since this $f(t)$ is real, but $x(t)$ is not, the general solution of (31) is the superposition of the real and imaginary parts of $x(t)$. It is also clear that $x(t) = O(1)$ but $x'(t) \neq O(1)$; in fact, neither the real nor the imaginary part of $x'(t)$ is $O(1)$. This proves more than what is claimed by the first part of (iv*).

In order to prove the second part of (iv*), it is sufficient to show that, if $f(t) = O(1)$ and $x(t) = O(1)$, then $x'(t) = O(1)$ holds by virtue of (31). But the two O -assumptions imply, by (30), that $x'' = O(1)$, and $x(t) = O(1)$ and $x''(t) = O(1)$ always imply that $x'(t) = O(1)$ (Hadamard).

9 bis. It may be mentioned that a trivial modification of the above proof of (iii) and (iv) leads to an easy proof of Weyl's central results ([3], pp. 223-231) concerning his *Grenzkreisfall* (results which replace (43) below by

$$(43^*) \quad f(t) - g(t) = \text{const.}$$

but, as seen from his proofs, apply under the more general assumption (43) also). In particular, the Corollary of the following theorem is nothing but [the obvious extension, from (43*) to (43), of] Weyl's main theorem ([3], p. 238) concerning his alternative of the *Grenzkreisfall* and the *Grenzpunktfall*:

(iv bis) *If $f(t)$ and $g(t)$ are (continuous, possibly complex-valued) functions for which*

$$(43) \quad f(t) - g(t) = O(1)$$

holds as $t \rightarrow \infty$, and if (31) has two linearly independent solutions, say $x = u(t)$ and $x = v(t)$, both of which are of class (L^2) , then there belongs

to every solution, $y(t)$, of (32), a pair of (continuous, bounded) functions, say $p(t)$ and $q(t)$, satisfying

$$y(t) = p(t)u(t) + q(t)v(t)$$

and

$$|p(t)|^2 + |q(t)|^2 \rightarrow C \text{ as } t \rightarrow \infty,$$

where the integration constant $C = C_{y(t)} \geq 0$ is 0 only when $y(t) \equiv 0$.

COROLLARY. The assumption (43) implies that all solutions of (31) are of class (L^2) if (and/or only if) all solutions of (32) are of class (L^2) ; simply because the last two formula lines imply that

$$y(t) = O(1)u(t) + O(1)v(t)$$

as $t \rightarrow \infty$.

In order to prove the Tauberian “ o -refinement,” (iv bis), of the latter “ O -theorem,” let the trivial solution, (35 bis), of (33) be excluded. Then (37) is applicable. But (37) and (43) imply that

$$\int_0^\infty |(\log r(t))'| dt \leq \text{const.} \int_0^\infty (|u(t)|^2 + |v(t)|^2) dt,$$

and the integral on the right of this inequality is convergent, since $u(t)$ and $v(t)$ are supposed to be of class (L^2) . Accordingly, $(\log r(t))'$ is absolutely integrable, hence integrable, over the half-line. In view of (36), the balance of the proof of (iv bis) is the same as that of (iii) or (iv) was.

Incidentally, (iv bis) can be thought of as a limiting case, $(\lambda; \lambda/(\lambda - 1)) = (1; \infty)$, of the following remark:

(iv* bis) If $f(t) - g(t)$ is of class (L^μ) , where $\mu > 1$, and if (31) has two linearly independent solutions, say $x = u(t)$ and $x = v(t)$, the squares of which are of class (L^λ) , where $\lambda = \mu/(\mu - 1)$, then there belong to every solution $y(t) \not\equiv 0$ of (32) a positive constant, C , and a pair of functions, $p(t)$ and $q(t)$, in terms of which $y(t)$ is representable in the form claimed by (iv bis).

In fact, the estimate (37) and the assumptions of (iv*) imply, by Hölder's inequality, the convergence of the integral on the right of the last formula line, and so (iv* bis) follows in the same way as (iv bis) did.

More useful than (iv* bis) is its limiting case $(\lambda; \lambda/(\lambda - 1)) = (\infty; 1)$:

(iv₀) If $f(t) - g(t)$ is of class $(L) = (L^1)$, and if (31) has two linearly independent solutions, say $x = u(t)$ and $x = v(t)$, which are $O(1)$

as $t \rightarrow \infty$, then there belong to every solution $y(t) \not\equiv 0$ of (32) a positive constant, C , and a pair of functions, $p(t)$ and $q(t)$, in terms of which $y(t)$ is representable in the form claimed by (iv bis).

In fact, this dual, $(\lambda; \mu) = (\infty; 1)$, of (iv bis), where $(\lambda; \mu) = (1; \infty)$, and even more than this dual, is contained in (iv), since, in view of the definition, (iv⁰), preceding (iv), the assumptions of the last theorem, (iv₀), are precisely those of (iv) itself.

10. After this deviation from the direction of (i), it will now be shown that the assertions of (i bis) need not hold if (3₀), the first of the assumptions of (i bis), is omitted; not even if (3), which then becomes a condition not implied by the remaining assumptions of (i bis), is satisfied. More than this is contained in the following negation:

(i* bis) *The convergence of both integrals (7) and the assumption (3) together do not prevent for (2) a solution $x(t) \not\equiv O(1)$.*

Conversely, the three assumptions of this negation, (i* bis), do not prevent for (2) a solution $x(t) \not\equiv 0$ which (instead of being, as in (i* bis), "large") is "small" as $t \rightarrow \infty$. In fact, the proof of (i* bis) will be such as to supply, for every $N > 0$, a $\phi(t)$ which satisfies the three assumptions of (i* bis) but is such that (2) admits of a solution $x(t) \not\equiv 0$ which is $O(t^{-N})$ as $t \rightarrow \infty$. The existence of functions $\phi(t)$ of either of these types can readily be concluded from the following rule of construction:

If $\chi = \chi(t)$ is any function having a continuous derivative, χ' , and if $\phi = \phi(t)$ denotes the (continuous) function

$$(44) \quad -\phi = \chi^2 \cos^2 t + \chi' \cos t - 3\chi \sin t,$$

then

$$(45) \quad x(t) = \exp \left(\int^t \chi(s) \cos s \, ds \right) \cos t$$

is a solution of

$$x'' + (1 + \phi(t))x = 0$$

(that is, of (2), where $\omega = 1$).

For similar purposes, this rule has been deduced in [4], pp. 394-395 [but with an error in sign, the sum of the second and third terms on the right of the above definition, (44), of ϕ , viz., the sum

$$-(\chi' \cos t - \chi \sin t) - 2\chi \sin t = \chi' \cos t - 3\chi \sin t,$$

having been given (loc. cit., middle of page 395, where χ is denoted by g), erroneously, as

$$(\chi' \cos t - \chi \sin t) + 2\chi \sin t = \chi' \cos t - \chi \sin t,$$

which, however, does not affect the construction of the possibilities described there].

Remark. If ϕ , rather than χ , is given, then (44), instead of being the definition of ϕ , is a (singular) *Riccati equation* for χ .

Let the above rule be applied to

$$(46) \quad \chi(t) = t^{-1} \cos t$$

(where, in order to exclude the singularity of (46) at $t=0$, the half-line is, e. g., $1 \leq t < \infty$). It is seen from (46) that $\chi'(t)$ is $-t^{-1} \sin t + O(t^{-2})$ as $t \rightarrow \infty$. Hence, from (44) and (46),

$$-\phi(t) = O(t^{-1})^2 - t^{-1} \sin t \cos t + O(t^{-2}) - 3t^{-1} \cos t \sin t,$$

which means that

$$(47) \quad \phi(t) = 2t^{-1} \sin 2t + O(t^{-2}).$$

On the other hand, since

$$\int_0^t s^{-1} \cos^2 s \, ds \sim \theta \log t \text{ as } t \rightarrow \infty,$$

where θ is a positive constant, it is seen from (46) that the solution (45) is of the form

$$(48) \quad x(t) = \exp(\theta \log t + o(\log t)) \cos t = t^{\theta+o(1)} \cos t,$$

hence $x(t) \neq O(1)$. Since, as shown by (47), both integrals (7) are convergent, and (3) is satisfied by the present $\phi(t)$, the proof of (i* bis) is complete.

It is also seen that, in order to obtain a $\phi(t)$ belonging to the arbitrary exponent, $-N$, mentioned after (i* bis), it is sufficient to multiply (46) by an arbitrary constant. In fact, (44) shows that (47) then becomes multiplied by a constant, and so all three conditions of (i* bis) remain satisfied. On the other hand, the absolute exponent, θ , occurring in (48) becomes replaced by $c\theta$, where c is an arbitrary constant; so that (45) is a solution $x(t) \neq 0$ which is $O(t^{-N})$ as $t \rightarrow \infty$, if c is chosen to be less than $-N/\theta$.

Appendix.

In view of (ii), the negation (i*), in which the assumption is the same as in (ii), is disappointing. In fact, one would like to have *some* asymptotic formula, that is, something like (4), rather than just the estimates (17), (19), even though just an (L^0) -condition on the coefficient function is made.

It turns out that such an asymptotic formula, having a structure similar to, but different from, that of (4), actually exists. In fact, the true theorem is as follows:

If $\phi(t)$ is of class (L^0) , then there belongs to every (real-valued) solution $x(t) \not\equiv 0$ of (2) a unique pair of positive integration constants, say a and α ($\leq 2\pi$), such that

$$x(t) = a \cos \left(\alpha + \int_0^t \{\omega^2 + \phi(s)\}^{\frac{1}{2}} ds \right) + \epsilon(t),$$

where the remainder term, $\epsilon(t) = \epsilon_{aa}(t)$, and its derivative satisfy (5).

This means that the (L^2) -assumption of (i) becomes superfluous if the assertion, (4), of (i) is replaced by the last formula line. The integral occurring in it is

$$\omega \int_0^t (1 + \phi(s)/\omega^2)^{\frac{1}{2}} ds = \omega \int_0^t (1 + \frac{1}{2}\phi(s)/\omega^2 - \dots) ds,$$

which, approximately, is

$$\omega \int_0^t (1 + \frac{1}{2}\phi(s)/\omega^2) ds = \omega t + \frac{1}{2} \int_0^t \phi(s) ds/\omega,$$

the corresponding phase in (4). But the error in the fluctuations, which is thus introduced by the neglect of the higher terms of the binomial expansion, can affect the asymptotic variation of the phase in such a fashion that the approximation becomes illegitimate without the (L^2) -assumption of (i). This makes (i*) understandable indeed.

Clearly, the theorem to be proved can be formulated as follows: If $\phi(t)$ is of class (L^0) , then there belongs to every solution, $x(t)$, of (2) a unique integration constant, say $c = a + ib$, satisfying

$$x(t) - cy(t) \rightarrow 0 \text{ and } x'(t) - cy'(t) \rightarrow 0 \text{ as } t \rightarrow 0,$$

where

$$y(t) = \exp i \int_0^t (\omega^2 + \phi(s))^{\frac{1}{2}} ds.$$

But, if (2) is identified with (31), then (ii₀) and (iii₀) show that the coefficient function, $f(t)$, of (31) is of class (*). It follows therefore from (iii) that the proof will be complete if it is shown that the function defined by the last formula line satisfies a differential equation, (32), whose coefficient function, $g(t)$, is such as to make the deviation $f(t) - g(t)$ a function of class (L^1).

In the deduction of this differential equation, it can, for the reasons explained in connection with (22 bis), be assumed that $\phi(t)$ has a continuous derivative, $\phi'(t)$. Then, since a differentiation of the last formula line gives

$$y' = i(\omega^2 + \phi)^{\frac{1}{2}}y,$$

one more differentiation shows that

$$y'' = \frac{1}{2}i(\omega^2 + \phi)^{-\frac{1}{2}}\phi'y - (\omega^2 + \phi)y.$$

This means that $y(t)$ becomes a solution of (32) if $g(t)$ is defined by

$$g(t) = \omega^2 + \phi(t) - \frac{1}{2}i(\omega^2 + \phi(t))^{-\frac{1}{2}}\phi'(t).$$

Since (31) is represented by (2), it follows that

$$\int_0^\infty |g(t) - f(t)| dt = \int_0^\infty |-\frac{1}{2}i(\omega^2 + \phi(t))^{-\frac{1}{2}}| |d\phi(t)|.$$

But $\phi(t)$ is supposed to be of class (L^0), that is, $\phi(t)$ satisfies (3₀) and (3). Consequently,

$$\int_0^\infty |g(t) - f(t)| dt \leq \int_0^\infty \text{const.} |d\phi(t)| < \infty.$$

Since this means that $g(t) - f(t)$ is of class (L^1), the proof is complete.

It is clear that, if $\omega(t)$ is a positive, continuous function which tends, as $t \rightarrow \infty$, to a positive limit, $\omega(\infty)$, then condition (3₀) is satisfied by the difference $\phi(t) = \omega(t) - \omega(\infty)$ if and only if it is satisfied by $\phi(t) = \omega^2(t) - \omega^2(\infty)$, the difference of the squares. Hence, the preceding theorem can be restated as follows:

If $\lambda(x)$, where $0 \leq x < \infty$, is a positive, continuous function satisfying

$$\int_0^\infty |d\lambda(x)| < \infty \text{ and } \lambda(\infty) > 0,$$

then there belongs to every solution, $\psi = \psi(x)$, of the corresponding wave equation,

$$d^2\psi/dx^2 + \{2\pi/\lambda(x)\}^2\psi = 0,$$

a unique integration constant, $c = a + ib$, such that the difference

$$\psi(x) - c \exp \left(2\pi i \int_0^x \{\lambda(s)\}^{-1} ds \right),$$

as well as the derivative of this difference, tends to 0 as $x \rightarrow \infty$.

In fact, the restrictions imposed on $\lambda(x)$ are equivalent to

$$\int_0^\infty |d\omega(x)| < \infty \text{ and } \omega(\infty) > 0,$$

where $\omega(x) = 2\pi/\lambda(x)$.

Applications of this theorem to the Hellinger decomposition of the spectral form of the differential equation (in case of a fixed boundary condition,

$$[\alpha\psi(x) + \beta d\psi(x)/dx]_{x=0} = 0, \quad \alpha^2 + \beta^2 > 0,$$

assigned for a given $x = 0$) and, in particular, to problems in wave mechanics (Kramers) will be given elsewhere.

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

-
- [1] A. Kneser, "Untersuchung und asymptotische Darstellung der Integrale gewisser linearer Differentialgleichungen bei grossen reellen Werthen des Arguments (Zweiter Aufsatz)," *Journal für die reine und angewandte Mathematik*, vol. 117 (1897), pp. 72-103.
 - [2] O. Perron, "Ueber ein vermeintliches Stabilitätskriterium," *Göttinger Nachrichten*, 1930, pp. 28-29.
 - [3] H. Weyl, "Ueber gewöhnliche Differentialgleichungen mit Singularitäten und die zugehörigen Entwicklungen willkürlicher Funktionen," *Mathematische Annalen*, vol. 68 (1910), pp. 222-269.
 - [4] A. Wintner, "The adiabatic linear oscillator," *American Journal of Mathematics*, vol. 68 (1946), pp. 385-397.

POLYNOMIAL LEAST SQUARE APPROXIMATIONS.*

By D. C. LEWIS.

1. Introduction. Suppose that we have a (not necessarily bounded) linear functional, whose domain is a class of sufficiently regular functions, $f(x)$, and whose range is the class of polynomials, $P_n(x)$, of degree $\leq n$. Suppose furthermore that this functional is such that $P_n(x) \equiv f(x)$ whenever $f(x)$ is itself a polynomial of degree $\leq n$. Functionals of this type are commonly used for obtaining polynomial "approximations," $P_n(x)$, to functions, $f(x)$. Elementary examples are afforded by the Lagrange interpolation polynomials and by the polynomials obtained by taking the first $n + 1$ terms of Taylor's series or the first $n + 1$ terms in an expansion in a series of orthogonal polynomials with respect to an arbitrary weight function. There are many other examples. We accordingly refer to the difference, $f(x) - P_n(x)$, as the "remainder."

The object of this paper is to obtain a simple explicit expression for this remainder in terms of the $(m + 1)$ -th derivative of $f(x)$ ($m \leq n$) and certain other elements independent of $f(x)$ but dependent upon the particular functional under consideration.

Our results are not applicable to the most general functional of the type described above but only to those which involve, in a certain general way, the principle of least squares, or, more generally, those expressible in the manner indicated by equation (3) below. Even so, our results are sufficiently general to present a unified approach to approximations of the Lagrange or Taylor type¹ and to those of the Legendre or Tchebichef type² as well as to many other previously uninvestigated types. After reading the paper, the reader will recognize that the Lagrange or Taylor type of approximation is obtained in the cases when $\alpha_0(x), \alpha_1(x), \dots, \alpha_p(x)$, introduced in the next section, are step functions with just the right discontinuities, totalling $n + 1$ in number, to determine $P_n(x)$ uniquely so as to minimize the quantity S of

* Received July 28, 1946.

¹ Cf. G. D. Birkhoff, "General mean value and remainder theorems with applications to mechanical differentiation and quadrature," *Transactions of the American Mathematical Society*, vol. 7 (1906), pp. 107-136.

² Cf. G. Szegő, *Orthogonal Polynomials*, American Mathematical Society Colloquium Publications, vol. 23 (1939).

equation (1), this minimum value being zero. On the other hand the Legendre or Tchebichef type of approximation is characterized by the fact that $\alpha_0(x)$ is a suitable function with an infinite number of points of increase while $p = 0$. The minimum value of S is usually *not* zero. The expression for the remainder is valid outside as well as inside the interval of orthogonality.

2. Specification of the least squares problem. Let $\alpha_0(x), \alpha_1(x), \alpha_2(x), \dots, \alpha_p(x)$ be $p + 1$ monotonic non-decreasing functions defined for $a \leq x \leq b$. Let $f(x)$ be a function of class C^{p-1} over an interval $A \leq x \leq B$, where $A \leq a < b \leq B$. We suppose also that the p -th derivative $f^{(p)}(x)$ exists almost everywhere with respect to α_p and is such that the Lebesgue-Stieltjes integral,

$$\int_a^b [f^{(p)}(x)]^2 d\alpha_p(x),$$

exists. A polynomial $P_n(x)$ of degree $\leq n$ whose coefficients are such as to render

$$(1) \quad S = \sum_{k=0}^p \int_a^b [f^{(k)}(x) - P_n^{(k)}(x)]^2 d\alpha_k(x)$$

a minimum, is said to be the polynomial of degree n which best fits the function $f(x)$. This "best fitting" is, of course, in the sense of least squares *with respect to* a given set of distributions $\alpha_0(x), \dots, \alpha_p(x)$, over the interval $a \leq x \leq b$. Since this is the only sense in which the expression is used in this paper, we omit in the sequel this detailed specification. Moreover it will be convenient to denote the relationship between a function $f(x)$ and its best fitting polynomial $P_n(x)$ in the following manner:

$$(2) \quad f(x) \sim P_n(x).$$

In general, $P_n(x)$ in (2) is uniquely determined by $f(x)$. This is well known to be the case if $\alpha_0(x)$ has at least $n + 1$ points of increase; but it is also true in many other circumstances: for example, when $\alpha_0(x), \alpha_1(x), \dots, \alpha_q(x)$ each have at least one point of increase, while α_{q+1} has at least $n - q$ points of increase. ($q + 1 \leq p$.) The general condition is that the quadratic form,

$$\sum_{k=0}^p \int_a^b [(d^k/dx^k)(a_0 + a_1x + a_2x^2 + \dots + a_nx^n)]^2 d\alpha_k(x),$$

in the $n + 1$ quantities a_0, a_1, \dots, a_n should be positive definite. Under this condition the reader will find without difficulty that the $n + 1$ linear

equations for the determination of the $n+1$ coefficients of $P_n(x)$ so as to minimize the expression S have a non-vanishing determinant. A further investigation of the solutions of these $n+1$ linear equations reveals the important fact that $P_n(x)$ can be written in the form,

$$(3) \quad P_n(x) = \sum_{k=0}^n \int_a^b K_{nk}(x, y) f^{(k)}(y) d\alpha_k(y),$$

where $K_{nk}(x, y)$ are polynomials in both x and y , independent of f . In the sequel we assume that $\alpha_0, \alpha_1, \dots, \alpha_p$ are such that $P_n(x)$ is thus always uniquely determined by $f(x)$. A simple but important consequence of this is

LEMMA 1. *If $f(x)$ is a polynomial of degree $\leq n$ and if $f(x) \sim P_n(x)$, then $f(x) \equiv P_n(x)$. In other words, by (3)*

$$(4) \quad Q_\mu(x) = \sum_{k=0}^p \int_a^b K_{nk}(x, y) Q_\mu^{(k)}(y) d\alpha_k(y),$$

for any polynomial $Q_\mu(x)$ of degree $\mu \leq n$.

The proof consists in the remark that the indicated determination of $P_n(x)$ gives S in (1) the value zero, which is clearly a minimum value for the non-negative S .

3. Statement and proof of the main theorem. From now on we assume that $f(x)$ is of class C^{m-1} ($m \geq p$) on the interval $A \leq x \leq B$ and that the m -th derivative exists almost everywhere, is of bounded variation, and is continuous at $x=b$. It must also be assumed that $f^{(m-1)}(x)$ is absolutely continuous. In case $m=p$, we assume that $f^{(m)}(x)$ and $\alpha_p(x)$ have no points of discontinuity in common.

Let $f(x) \sim P_n(x)$. We now state our principal

THEOREM. *There exists a function $G_n^m(x, t)$, independent of $f(x)$, with $p \leq m \leq n$, such that*

$$(5) \quad f(x) = P_n(x) + (1/m!) \int_a^b G_n^m(x, t) df^{(m)}(t) + (1/m!) \int_b^x (x-t)^m df^{(m)}(t)$$

for any point x on the interval $A \leq x \leq B$. Moreover the function $G_n^m(x, t)$ is explicitly given for each value of t as that polynomial in x of degree $\leq n$ which best fits the function $g_m(x, t)$ defined as follows:

$$\begin{aligned} g_m(x, t) &= (x-t)^m \text{ if } x \leq t \\ g_m(x, t) &= 0 \quad \text{if } x > t. \end{aligned}$$

We begin our proof by referring to Taylor's theorem with the integral form of the remainder:

$$(6) \quad f(x) = \sum_{h=0}^m (f^{(h)}(b)/h!) (x-b)^h + (1/m!) \int_b^x (x-t)^m df^{(m)}(t).$$

This shows that the identity (5) to be established is equivalent to the following:

$$(7) \quad P_n(x) = \sum_{h=0}^m (f^{(h)}(b)/h!) (x-b)^h - (1/m!) \int_a^b G_n^m(x, t) df^{(m)}(t).$$

To establish (7), we insert the right hand member of (6), after changing x to y , into (3). If we use Lemma 1, we discover at once that

$$\begin{aligned} P_n(x) &= \sum_{h=0}^m (f^{(h)}(b)/h!) (x-b)^h \\ &\quad + \sum_{k=0}^n \int_a^b K_{nk}(x, y) (1/(m-k)!) \int_b^y (y-t)^{m-k} df^{(m)}(t) d\alpha_k(y). \end{aligned}$$

If we reverse the order of the iterated integrals, we obtain

$$\begin{aligned} P_n(x) &= \sum_{h=0}^m (f^{(h)}(b)/h!) (x-b)^h \\ &\quad + \sum_{k=0}^p (1/(m-k)!) \int_b^a \int_a^t K_{nk}(x, y) (y-t)^{m-k} d\alpha_k(y) df^{(m)}(t). \end{aligned}$$

But this reduces at once to (7), if we make use of the formulas,

$$\begin{aligned} G_n^m(x, t) &= \sum_{k=0}^p \int_a^b K_{nk}(x, y) g_m^{(k)}(y, t) d\alpha_k(y) \\ &= \sum_{k=0}^p (m!/(m-k)!) \int_a^t K_{nk}(x, y) (y-t)^{m-k} d\alpha_k(y), \end{aligned}$$

which are clear consequences of the definition of G_n^m and of (3).

This completes the formal proof. The only delicate step of rigor is the reversal of the order of the iterated Stieltjes integrals, which, of course, involves Fubini's theorem.³ The application of Fubini's theorem to the double integral,

$$\int \int K_{nk}(x, y) g_m^{(k)}(y, t) d\alpha_k(y) df^{(m)}(t),$$

over the region, $a \leq y \leq b$, $a \leq t \leq b$, is, however, immediate when the integrand is continuous. The only case of discontinuity occurs when $k = m = p$, in which case $g_m^{(k)}(y, t)$ is not defined at the points of discontinuity, namely the points for which $y = t$. In virtue of the hypothesis

³ Cf. Stanislaw Saks, *Theory of the Integral* (2nd edition), p. 77.

that $\alpha_p(x)$ and $f^{(m)}(x)$ have no common points of discontinuity, it is easily seen that the set $y = t$ has measure zero with respect to the two dimensional measure function, $m(E) = \int_E \int d\alpha_p(y) df^{(m)}(t)$. Hence, Fubini's theorem can be applied even in this case.

The reason for the hypothesis that $f^{(m)}(x)$ be continuous at $x = b$ and also for the hypothesis that $f^{(m-1)}(x)$ be absolutely continuous comes in the proof of Taylor's Theorem in the form indicated in (6). This formula is established by repeated integration by parts in a well known manner, but the proof requires the assumption that $f^{(m)}(b-0) = f^{(m)}(b)$, if $b < x$, and that $f^{(m)}(b+0) = f^{(m)}(b)$, if $b > x$.⁴ It is equally necessary to assume that $f^{(m-1)}(x)$ is an indefinite integral of $f^{(m)}(x)$, which means that $f^{(m-1)}(x)$ must be absolutely continuous. That these conditions are not only required by the proof but are also logically necessary for the validity of (6) may be shown easily with the help of simple counter examples.

4. Interpretation of the main theorem in case $a \leq x \leq b$. If $a \leq x \leq b$, it is obvious from the definition of $g_m(x, t)$ that (5) can be written in the form,

$$(8) \quad f(x) - P_n(x) = (1/m!) \int_a^b [G_n^m(x, t) - g_m(x, t)] df^{(m)}(t).$$

Now, in its dependence upon x , we have $g_m(x, t) \sim G_n^m(x, t)$. Equation (8) thus indicates the manner in which the remainder for the *general* function $f(x)$ is synthesized from the remainder for the *special* function $g_m(x, t)$ and from the values of $f^{(m)}(x)$.

The result embodied in (8) is probably well known in various special cases, although specific references are unknown to the author.

In the fairly general situation where at least one of the α 's has infinitely many points of increase, one can, with fixed α 's and fixed m , let $n \rightarrow \infty$. It is then clear from (8) that, if G_n^m is uniformly bounded and approaches $g_m(x, t)$ almost everywhere with respect to $\int |df^{(m)}(t)|$, $P_n(x)$ must approach $f(x)$. Here, of course, m may be taken to be any number $\geq p$. In case $p = 0$, m may also be taken to be zero, and then $f^{(0)}(x)$, is, of course, $f(x)$ itself. These remarks have obvious connections with well known results on the convergence of series of orthogonal polynomials.

⁴ Cf. Saks, *loc. cit.*, p. 102.

5. Taylor's remainder theorem as a special case of the main theorem.

Although our main theorem is a consequence of Taylor's theorem with the integral remainder, it is of interest to show that the converse is also true. This will give an indication of the scope of our result by discussing a case not covered in the last paragraph of the preceding section.

Let $p = m = n$. Let c be any point of the interval (a, b) where $f^{(m)}(x)$ is continuous, and let $\alpha_k(x) = 0$ for $a \leq x \leq c$, and $\alpha_k(x) = 1$ for $c < x \leq b$, $k = 0, 1, 2, \dots, n$. Then the requirement that S be a minimum reduces to the following:

$$P_n^{(k)}(c) = f^{(k)}(c), \text{ for } k = 0, 1, 2, \dots, n.$$

Hence $P_n(x)$ is the sum of the first $n + 1$ terms of the Taylor expansion of $f(x)$ about the point $x = c$. Similarly, since $g_n(x, t) \sim G_n^n(x, t)$, we have

$$\frac{d^k G_n^n(c, t)}{dx^k} = \frac{d^k g_n(c, t)}{dx^k}, \quad k = 0, 1, 2, \dots, n.$$

This shows that $G_n^n(x, t) = (x - t)^n$, if $c < t$,

$$G_n^n(x, t) = 0, \text{ if } t < c.$$

Equation (5) can therefore in this particular instance be written as follows:

$$\begin{aligned} f(x) &= P_n(x) + (1/n!) \int_c^b (x - t)^n df^{(n)}(t) \\ &\quad + (1/n!) \int_b^x (x - t)^n df^{(n)}(t) \\ &= \sum_{k=0}^n (f^{(k)}(c)/k!) (x - c)^k + (1/n!) \int_c^x (x - t)^n df^{(n)}(t), \end{aligned}$$

which is merely equation (6) with b replaced by c .

UNIVERSITY OF MARYLAND.

A NOTE ON LAPLACE TRANSFORMS OF FUNCTIONS WHOSE SPECTRA ARE CONFINED TO A GIVEN SET.*

By E. K. HAVILAND.

It is known that a necessary and sufficient condition that a function $f(x)$ should be representable in the form

$$\int_0^{\infty} e^{-xt} d\beta(t),$$

where $\beta(t)$ is bounded and non-decreasing and the integral converges for $0 \leq x < +\infty$, is that $f(x)$ should be completely monotonic in $0 \leq x < +\infty$. If the requirements on $\beta(t)$ are merely that it be non-decreasing and that the integral converge for $0 < x < +\infty$, the corresponding necessary and sufficient condition is that $f(x)$ should be completely monotonic in $0 < x < +\infty$. For these theorems, several proofs are known, of which the earliest appears to be that which is an immediate consequence of Hausdorff's solution of the continuous momentum problem.¹

The purpose of the present note is to extend the foregoing results on the Laplace transform to the case where the spectrum of the function $\beta(t)$ is confined to a preassigned subset Γ of the interval $[0, +\infty)$, the extension being based on results previously obtained by the present author in connection with the momentum problem.²

By the spectrum of the function $\beta(t)$ is meant the set of points τ such that $\beta(t_2) - \beta(t_1) > 0$ for every interval (t_1, t_2) such that $t_1 < \tau < t_2$.

Furthermore, we recall that a function $f(x)$ is defined to be completely monotonic in an interval $(0, +\infty)$, if its p -th derivative satisfies the condition

$$(-1)^p f^{(p)}(x) \geq 0, \quad (p = 0, 1, 2, \dots),$$

in the interior of the interval and if the function $f(x)$ is continuous at the

* Received November 27, 1946.

¹ F. Hausdorff, "Summationsmethoden und Momentfolgen. II," *Mathematische Zeitschrift*, vol. 9 (1921), pp. 280-299, especially, pp. 282-287.

² E. K. Haviland, "On the momentum problem for distribution functions in more than one dimension," *American Journal of Mathematics*, vol. 58 (1936), pp. 164-168.

end point 0, if the latter is included in the interval. This is equivalent to the infinitely many difference equations ³

$$(-1)^p \begin{pmatrix} f(t_0) & f(t_1) & \cdots & f(t_p) \\ t_0 & t_1 & \cdots & t_p \end{pmatrix} \geq 0,$$

where the second factor denotes the p -th divided difference of $f(t)$ formed for any points $t_0, t_1, \dots, t_p (\geq 0)$.

Our principal result is then given by

THEOREM I. *A necessary and sufficient condition that a function $f(x)$ be representable in the form*

$$(1) \quad f(x) = \int_0^\infty e^{-xu} d\beta(u),$$

where $\beta(u)$ is bounded and non-decreasing and has its spectrum confined to the sub-set Γ of $[0, +\infty)$ and the integral converges for $0 \leq x < +\infty$, is that $f(x)$ be completely monotonic in $[0, +\infty)$ and that, if $a_0 + a_1 e^{-u} + \cdots + a_n e^{-nu}$ be any exponential polynomial non-negative on Γ , then the functional value

$$(2) \quad a_0 f(x) + a_1 f(x+1) + \cdots + a_n f(x+n) \geq 0$$

for all $x \geq 0$.

Proof. If in (1) we make the transformation $e^{-u} = t$ or $u = -\log t$, (u real for $t > 0$), and if we denote the bounded non-decreasing function $-\beta(-\log t)$ by $\alpha(t)$, we may write $f(x)$ in the form

$$(3) \quad f(x) = \int_0^1 t^x d\alpha(t),$$

in which case the spectrum of $\alpha(t)$ is confined to a set C contained in the interval $[0, 1]$, C being the image on the t -axis of the set Γ on the u -axis. Then, if $\phi(t) \geq 0$ is integrable over C with respect to α ,

$$\Phi(x) = \int_0^1 t^x \phi(t) d\alpha(t)$$

exists for all $x \geq 0$ and is a continuous function of x there, except perhaps at $x = 0$, at which point a possible right-hand discontinuity can always be

³ Cf. F. Hausdorff, *ibid.*, p. 284.

avoided by defining ⁴ $\alpha(0) = \alpha(+0)$, a procedure which, of course, imposes a restriction on $\Phi(0)$.

Moreover, $\Phi(x) \geq 0$ for all x , ($0 \leq x < +\infty$), and

$$(-1)^k \Phi^{(k)}(x) = (-1)^k \int_0^1 t^x (\log t)^k \phi(t) d\alpha(t)$$

exists and is ≥ 0 for all $x > 0$, ($k = 1, 2, \dots$), so that $\Phi(x)$ is completely monotonic in $x \geq 0$. In particular, this is true of any polynomial

$$(4) \quad \phi(t) = a_0 + a_1 t + \dots + a_n t^n,$$

non-negative on C , in which case,

$$(5) \quad \begin{aligned} \Phi(x) &= \int_0^1 t^x \{a_0 + a_1 t + \dots + a_n t^n\} d\alpha(t) \\ &= a_0 f(x) + a_1 f(x+1) + \dots + a_n f(x+n) \geq 0. \end{aligned}$$

This completes the proof of the necessity of the conditions stated in Theorem I, and it may be noted that the functional (2) is not only non-negative but completely monotonic.

Suppose, conversely, that the function $f(x)$ is completely monotonic for $x \geq 0$ and that, for every polynomial (4) and every $x \geq 0$ (or, at least, for $x=0$, which is all that is needed for the present proof), (2) holds. Let $x=0$ in (2) and put $f(n) = \mu_n$, ($n=0, 1, 2, \dots$). Then to every polynomial (4) non-negative on C there corresponds the non-negative functional value

$$a_0 \mu_0 + a_1 \mu_1 + \dots + a_n \mu_n.$$

This condition is known ⁵ to be sufficient for the existence of a bounded non-decreasing function, $\alpha(t)$, whose spectrum is contained in C and which is such that $\mu_n = \int_0^1 t^n d\alpha(t)$.

Since the corresponding moment function $\mu(x) = \int_0^1 t^x d\alpha(t)$ is completely monotonic for $x \geq 0$, the $\mu_n = \mu(n)$ form a completely monotonic sequence and, as $\Sigma(t_n)^{-1} = \Sigma n^{-1}$ diverges, this completely monotonic sequence

⁴ Cf. F. Hausdorff, *ibid.*, p. 286.

⁵ E. K. Haviland, *ibid.*, p. 164.

determines uniquely^o a completely monotonic function $f(x)$ for which $f(n) = \mu_n$. Consequently, our original function $f(x)$, which was completely monotonic in $x \geq 0$ by hypothesis, must be identical with $\mu(x)$, i. e., on setting $t = e^{-u}$ and $\beta(u) = -\alpha(e^{-u})$,

$$f(x) = \int_0^\infty e^{-xu} d\beta(u),$$

where $\beta(u)$ is a bounded monotonic function whose spectrum is contained in the preassigned set Γ .

In a similar way, we have

THEOREM II. *A necessary and sufficient condition that a function $f(x)$ be representable in the form (1), where $\beta(u)$ is non-decreasing and has its spectrum confined to the sub-set Γ of $[0, +\infty)$ and the integral converges for $0 < x < +\infty$, is that $f(x)$ be completely monotonic in $0 < x < +\infty$ and that, if $a_0 + a_1 e^{-u} + \dots + a_n e^{-nu}$ be any exponential polynomial non-negative on Γ , then the functional value (2) be non-negative for all $x > 0$.*

In fact, the necessity of the condition follows exactly as before. As to the sufficiency, we observe that, for any fixed $\xi > 0$, it follows as in the proof of Theorem I that

$$f(\xi + n) = \mu'_n = \int_0^1 t^n d\alpha_\xi(t), \quad (n = 0, 1, \dots),$$

uniquely determines the function

$$f(\xi + x) = \mu_\xi(x) = \int_0^1 t^x d\alpha_\xi(t),$$

which is completely monotonic in $x \geq 0$, $\alpha_\xi(t)$ being a bounded non-decreasing function whose spectrum is confined to the set G . On replacing x by $x - \xi$, we obtain

$$f(n) = \int_0^1 t^n \cdot t^{-\xi} d\alpha_\xi(t) = \int_0^1 t^n d\alpha(t), \quad (n = 0, 1, 2, \dots)$$

where

$$\alpha(t) = \int_0^t \tau^{-\xi} d\alpha_\xi(\tau),$$

so that, by virtue of the uniqueness of the solution of this moment problem, $t^{-\xi} d\alpha_\xi(t) = d\alpha(t)$, for all $\xi > 0$. Finally,

$$f(x) = \int_0^1 t^x d\alpha(t)$$

^o Cf. F. Hausdorff, *ibid.*, p. 284.

for $x > 0$. As the spectrum of $\alpha_\xi(t)$ is confined to the set C for all $\xi > 0$, the same will be true of $\alpha(t)$, save possibly for the addition of the point $t = 0$. But, if we define $\alpha(0) = \alpha(+0)$, where $\alpha(+0)$ may be $-\infty$, this situation will cause no difficulty, inasmuch as a spectrum is necessarily closed. This completes the proof of Theorem II.

We next consider sub-sets Γ of $[0, +\infty)$ for which there can be found a basis of polynomials non-negative on Γ . In particular, let Γ be $[0, 1]$, to which corresponds the interval $C : [e^{-1}, 1]$. Since any polynomial non-negative on the latter interval can be represented (at least in the limit) in the form

$$\sum A(1-x)^m(x-e^{-1})^n,$$

where $m, n = 0, 1, 2, \dots$ and the A 's are non-negative,⁷ it follows that (4) may here be replaced by

$$\begin{aligned} (1-t)^m(t-e^{-1})^n &= e^{-n}(-1)^n(1-t)^m(1-et)^n \\ &= e^{-n}(-1)^n \sum_{j=0}^m C_j^m(-1)^j \sum_{\nu=0}^n C_\nu^n(-1)^\nu e^\nu t^{j+\nu}, \end{aligned}$$

the C_j^m, C_ν^n being binomial coefficients.

If we neglect the positive factor e^{-n} , the analogue of equation (2) is now

$$(6) \quad \sum_{j=0}^m C_j^m(-1)^j \left\{ (-1)^n \sum_{\nu=0}^n C_\nu^n(-1)^\nu e^\nu f(x+j+\nu) \right\} \geq 0,$$

or

$$(6 \text{ bis}) \quad (-1)^m \Delta_j^m \{ \Delta_\nu^n \{ e^\nu f(x+j+\nu) \} \} \geq 0,$$

where $m, n = 0, 1, 2, \dots$ and $x \geq 0$ in the case of Theorem I or $x > 0$ in the case of Theorem II. The latter, for instance, may be stated explicitly as

THEOREM III. *Necessary and sufficient in order that a function $f(x)$ be representable in the form (1), where $\beta(u)$ is a non-decreasing function with spectrum confined to $[0, 1]$ is that*

- (i) *the function $f(x)$ be completely monotonic in $x > 0$,*
- (ii) *the condition (6) or (6 bis) hold for all $x > 0$.*

⁷ Cf., e. g., Pólya und Szegő, *Aufgaben und Lehrsätze aus der Analysis*, vol. II, p. 83, Ex. 49.

The infinitely many difference conditions contained in (6) suggest that possibly (ii) implies (i), the more so because, as is well known,⁸

$$(7) \quad \Delta_k^n f(a) = (n!)^{-1} f^{(n)}(\xi), \quad a < \xi < a + n.$$

That (ii), however, does not imply (i) can be seen from the example

$$f(x) = 2 + \sin(2\pi x),$$

which is positive for all x and has the period 1. On interchanging the order of the summations in (6), we obtain

$$(8) \quad (-1)^{m+n} \sum_{\nu=0}^n C_\nu^n (-1)^\nu e^\nu \{ \Delta_j^m f(x + \nu + j) \}.$$

Due to the periodicity of $f(x)$, (8) vanishes for all x if $m \geq 1$, while, if $m = 0$, it reduces to

$$f(x) (-1)^n \sum_{\nu=0}^n C_\nu^n (-e)^\nu = f(x) (e-1)^n > 0,$$

so that in any case condition (ii) is fulfilled. In spite of this, $f(x)$ is not even monotone and hence (i) is not satisfied.

P. Hartman has pointed out to me that condition (ii) of Theorem II, which, in virtue of (7), may be written as

$$(-1)^m \frac{d^n}{dx^n} \{ e^x f^{(m)}(x) \} \geq 0,$$

can be replaced by the simpler conditions

$$(ii^*) \quad f(x) + f'(x) \text{ is completely monotone for } x > 0,$$

or

$$(ii^{**}) \quad \text{the sequence of numbers } f(x), f'(x), f''(x), \dots \text{ is bounded for some fixed } x = x_0 > 0.$$

The necessity of these conditions is seen directly, if it be observed that the upper limit of integration in (1) is now 1. As to the sufficiency, (ii*) implies that the sequence $\{|f^{(n)}(x)|\}$, where $f^{(n)} = d^n f/dx^n$, is monotone non-increasing and hence (ii**) is valid. But, if $\beta(t)$ had a point $x > 1$ in its spectrum,

$$(-1)^n f^{(n)}(x) = \int_0^\infty e^{-xt} t^n d\beta(t) \geq (\gamma - \delta)^n \int_{\gamma-\delta}^\infty e^{-xt} d\beta(t),$$

⁸ Cf. O. Hölder, "Zur Theorie der trigonometrischen Reihen," *Mathematische Annalen*, vol. 24 (1884), p. 183.

and as δ may be so chosen that $\gamma - \delta > 1$, the last expression would become infinite with n , thus proving the sufficiency of (ii**).

Moreover, Theorem III remains true, if the interval $[0, 1]$ is replaced by any interval $[a, b]$, where $0 \leq a < b < \infty$, and if the set of conditions (i)-(ii) is replaced by either one of the following sets:

(i*) $f(x)e^{ax}$ is completely monotonic for $x > 0$,

(ii*) $bf(x) + f'(x)$ is completely monotonic for $x > 0$

or

(i**) $f(x)$ is completely monotonic for $x > 0$,

(ii**) $f(x) = O(e^{-ax})$, $x \rightarrow \infty$,

(iii**) $f^{(n)}(x) = O(b^n)$, $n \rightarrow \infty$, for some fixed $x = x_0 > 0$.

THE JOHNS HOPKINS UNIVERSITY.

ON THE REPRESENTATIONS OF THE SYMMETRIC GROUP.*

(Second Paper)

By G. DE B. ROBINSON.

Introduction. The close relationship between the theory of the representations of the full linear group L and that of the symmetric group was first remarked by Schur and has been studied by many authors.¹ In particular, the reduction of the direct product of two irreducible representations of L is given in a theorem of Littlewood and Richardson proved by the author in 1938.² Since the substance of the present paper is so closely related to that of SG_1 it has been given the same title and designated 'Second Paper.'

The paper is divided into two parts. In Part I the familiar Young diagram³ is generalized in the following manner. Consider an irreducible representation $[\alpha]$ of the symmetric group S_l and a representation $[\beta]$ of S_m where $l = m + n$ and $\beta_i \leq \alpha_i$ for all i . The *right* diagram $[\beta]$ is said to be *contained* in $[\alpha]$, and the symbolic difference $[\alpha] - [\beta]$ is taken to represent the *skew diagram* composed of these nodes of $[\alpha]$ not belonging to $[\beta]$. Just as one speaks of a *standard* right diagram so one can speak of a *standard* skew diagram, and these latter lead to a representation of S_n which is reducible. This representation is called a *skew representation* as compared with the familiar right representation which is irreducible. The irreducible components of $[\alpha] - [\beta]$ are determined in §3, and some formulae connecting the degrees are developed in §4.

In Part II this theory is applied to give a new proof of the Murnaghan-Nakayama⁴ recursion formula.

The chief purpose of developing the theory in Part I has been to relate the two remarkable papers by Nakayama⁴ to the main body of Young's work.

* Received June 22, 1946.

¹ For references consult [5]. The recent paper [1] by Stig Com  t is of interest, and complete accounts of the background will be found in Littlewood [2] and Weyl [7].

² [5]. This paper will be referred to as SG_1 .

³ Young consistently used the word *tableau*. Several recent writers have since favoured the term *diagram* to which usage we conform. It is necessary to have a term for the intersection of a row and column in a diagram. Young spoke of 'place' or 'position,' but *node* seems to be convenient and suggestive.

⁴ [4], Part I, p. 107. This paper will be referred to as N_1 , Part II as N_2 .

The tools developed, particularly formulae 4.3, 4.4 and 4.5, have proved useful and it is hoped to publish the results of their application to the modular theory shortly.

PART I.

A generalization of the Young diagram.

1. Following the notation of SG_1 , let us denote by $P_n(L)$ the n -th power representation of the full linear group L and write

$$1.1 \quad \Delta_{(a)}(L) = P_{a_1}(L) \times P_{a_2}(L) \times \cdots \times P_{a_h}(L).$$

Since $P_n(L) = \{n\}$, we have also

$$1.2 \quad \Delta_{(a)}(L) = \{\alpha_1\} \times \{\alpha_2\} \times \cdots \times \{\alpha_h\},$$

and in general we may consider the direct product of two such representations

$$1.3 \quad \Delta_{(\alpha)}(L) = \Delta_{(\beta)}(L) \times \Delta_{(\gamma)}(L),$$

where $\Sigma\beta_i = m$, $\Sigma\gamma_i = n$ and $\Sigma\alpha_i = m + n = l$.

In SG_1 the chief interest was in proving Littlewood and Richardson's theorem for the reduction of $\{\beta\} \times \{\gamma\}$. Here we consider the more general problem of the reduction of the direct product 1.3 when $[\alpha] = [1^l]$:

$$1.4 \quad \begin{aligned} \Delta_{(1^l)}(L) &= \Delta_{(1^m)}(L) \times \Delta_{(1^n)}(L) \\ &= \Sigma\{\beta\} \times \{\gamma\} = \Sigma\{\alpha\}. \end{aligned}$$

How many times will a given irreducible component $\{\alpha\}$ appear in 1.4? Certainly, its coefficient will receive contributions from many $\{\beta\} \times \{\gamma\}$. To find such contributions one has recourse to the Littlewood and Richardson rule which we restate here as follows:

1.5 LR_1 : Take the diagram $[\beta]$ intact and add to it the symbols of the first row of $[\gamma]$. These may be added to one row of $[\beta]$, or the symbols may be divided without disturbing their order into any number of sets, the first set being added to one row of $[\beta]$, the second to a subsequent row, and so on. After the addition no row must contain more symbols than a preceding row, and no two added symbols may appear in the same column.

Next add the second row of $[\gamma]$ according to the same rules,

followed by the remaining rows in succession until all the symbols of $[\gamma]$ have been used.

LR_2 : These additions shall be such that each symbol of a given row of $[\gamma]$ in the compound diagram must appear in a later row than the symbol on the same column from a preceding row of $[\gamma]$.

A necessary condition for $\{\alpha\}$ to be an irreducible component of $\{\beta\} \times \{\gamma\}$ is that the diagram $[\beta]$ be *contained in* the diagram $[\alpha]$ in the sense that $\beta_i \leq \alpha_i$ for all i . Since

$$1.6 \quad \{\beta\} \times \{\gamma\} = \{\gamma\} \times \{\beta\},$$

the same holds for $[\gamma]$.

Since $\{\beta\}$ appears in $\Delta_{(1^m)}(L)$ with frequency f_β , the frequency of appearance of $\{\alpha\}$ on the right hand side of 1.4 is given by

$$\sum_{\beta} f_{\beta} \sum_{\gamma} a\lambda_{\beta}^{\gamma} f_{\gamma};$$

$a\lambda_{\beta}^{\gamma}$ is a positive integer, or zero, and gives the number of ways in which the diagram $[\gamma]$ can be built on the diagram $[\beta]$ to yield the diagram $[\alpha]$. On the other hand $\{\alpha\}$ appears in $\Delta_{(1^l)}(L)$ with a frequency f_{α} , so that

$$1.7 \quad f_{\alpha} = \sum_{\beta} f_{\beta} \sum_{\gamma} a\lambda_{\beta}^{\gamma} f_{\gamma} = \sum_{\gamma} f_{\gamma} \sum_{\beta} a\lambda_{\gamma}^{\beta} f_{\beta}.$$

It follows from 1.6 that

$$1.8 \quad a\lambda_{\gamma}^{\beta} = a\lambda_{\beta}^{\gamma},$$

and we may write 1.7 in the form

$$1.9 \quad f_{\alpha} = \sum_{\beta} \phi_{\alpha}^{\beta} f_{\beta},$$

where

$$1.10 \quad \phi_{\alpha}^{\beta} = \sum_{\gamma} a\lambda_{\beta}^{\gamma} f_{\gamma}.$$

If we denote the diagram conjugate to $[\alpha]$ by $[\alpha^*]$ then a little consideration will show that

$$1.11 \quad a\lambda_{\gamma}^{\alpha\beta^*} = a\lambda_{\gamma}^{\beta}.$$

In the next section we shall see how the formulae 1.9 and 1.10 can be obtained from a different point of view.

2. In studying a representation ω of a finite group G it is fruitful to limit attention to a sub-group H of G and investigate the reducibility of ω with regard to H . Let us apply this procedure to S_l and, since our knowledge of the symmetric group is relatively so complete, let us consider the reducibility of a representation $[\alpha]$ of S_l with regard to a sub-group S_m , where $l = m + n$. In effect we restrict attention to permutations of S_l of the form

$$2.1 \quad (1, 2, \dots, m) (m+1, \dots, l).$$

Such permutations form a sub-group of S_l which is the direct product of S_m and S_n . Let us suppose that the corresponding standard diagram $[\alpha]$ is of the form:

$$\begin{array}{ccccccc}
a_{11} & \cdots & a_{1\beta_1} & a_{1,\beta_1+1} & \cdots & a_{1\alpha_1} \\
a_{21} & \cdots & a_{2\beta_2} & a_{2,\beta_2+1} & \cdots & a_{2\alpha_2} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
a_{h1} & \cdots & a_{h\alpha_h}
\end{array}$$

From the restriction that the symbols must appear in the assigned or 'natural' order in any row or column of a standard diagram, it follows that the first m symbols must form a standard sub-diagram $[\beta]$ contained in $[\alpha]$. The remaining n symbols may be thought of as forming a standard *skew diagram*, which we may denote $[\alpha] \rightarrow [\beta]$. It will be understood that this symbol has a meaning only if $[\beta]$ is contained in $[\alpha]$.

Young's rule for constructing the actual matrix representing the transposition $(a_r a_{r+1})$ in the orthogonal representation $[\alpha]$ of S_t is as follows:⁵

2.3 Put (i) unity in the leading diagonal where the associated standard diagram has a_r, a_{r+1} in the same row;

(ii) -1 in the leading diagonal where the associated standard diagram has a_r, a_{r+1} in the same column;

(iii) a quadratic matrix

$$\begin{bmatrix} -\rho & \sqrt{1-\rho^2} \\ \sqrt{1-\rho^2} & \rho \end{bmatrix},$$

⁵ [5] reference Young [21], Part VI, pp. 217-8. Cf. also Thrall [6].

the elements being in the positions of intersections of a pair of rows and columns whose diagrams differ only by the exchange of a_r and a_{r+1} .

The quantity ρ is defined by the relation

$$\rho^{-1} = (\mu - \lambda) + (\gamma - \delta),$$

where $a_r = a_{\mu\delta}$ and $a_{r+1} = a_{\lambda\gamma}$, and we may suppose that $\mu > \lambda$ in the standard diagram in question. Young calls ρ the *axial projection* of $a_r a_{r+1}$ in $[\alpha]$.

It is important that ρ is defined in terms of the positions of the two symbols a_r, a_{r+1} only in $[\alpha]$. Consequently, the matrices representing the permutations of S_m are in reduced form as they stand and have as irreducible components all possible $[\beta]$ contained in $[\alpha]$. For exactly the same reason the matrices provide a representation of S_n , the variables corresponding to the ordinary standard diagrams being the standard skew diagrams. We can think of the formula 1.9 as representing the reduction of the direct product of S_m and S_n , where ϕ_a^β is not only the frequency with which $[\beta]$ appears in $[\alpha]$ when S_i is restricted to S_m , but also the *degree* of a reducible representation of S_n , which may be denoted $[\alpha] - [\beta]$ and called a *skew representation*.

The skew diagram can be characterized as *standard* in just the same way as a right diagram, as already remarked. We have in fact a generalization which coincides with the right diagram under certain obvious conditions; e.g. in the case where $[\alpha] = [5, 4]$ and $[\beta] = [2^2]$.

3. The problem of determining the integers $a_{\lambda\beta}^\gamma$ is of considerable interest. What it amounts to is to establish a correlation between the two interpretations of the preceding paragraphs. We have on the one hand the ϕ_a^β standard skew diagrams and we wish to associate them with the different direct products $\{\beta\} \times \{\gamma\}$ which yield contributions to the coefficients of $\{\alpha\}$ in 1.4. As in SG₁ § 5, MacMahon's lattice permutations⁶ provide the clue.

Consider a permutation of the n letters

$$3.1 \quad c_1^{\gamma_1} c_2^{\gamma_2} \cdots c_h^{\gamma_h},$$

where $\Sigma \gamma_i = n$. Such a permutation is said to be *lattice* if amongst the first r terms of it the number of c_1 's \geq the number of c_2 's $\geq \cdots \geq$ the number of c_h 's, for all r . If we add a second suffix to the c_i 's according to the order

⁶ Cf. also [2], pp. 67-71 and p. 95.

of their appearance for each i we may define the indices of the permutation. Considering first only the c_1 's and the c_2 's, if c_{2s} follows c_{1t} and precedes $c_{1,t+1}$ its index is defined to be

$$3.2 \quad i_{12s} = s - t,$$

which may be positive, zero or negative. The resulting permutation of $c_1^{\gamma_1} c_2^{\gamma_2}$ is lattice if and only if no $i_{12s} > 0$. Similarly, we may define indices i_{23s} , i_{34s} , etc., and a permutation of the letters 3.1 is lattice if and only if no $i_{x,x+1,s} > 0$ for $x = 1, 2, \dots, h - 1$.

The important fact in this connection is that a non-lattice permutation may be made lattice according to a specific sequence of changes. This step-by-step process is as follows:

(a) Considering only the c_1 's and the c_2 's in the permutation, take the first c_2 with the greatest positive i_{12s} and change it into a c_1 . Reallocating the second suffixes, repeat the process, continuing until all the c_1 's and the c_2 's are lattice.

(b) Considering only the c_2 's and the c_3 's in the permutation so modified, take the first c_3 with greatest positive i_{23s} and change it into a c_2 . If this change upsets the 1-2 lattice property, correct for it by changing a c_2 into a c_1 according to (a); this may or may not be the new c_2 . Reallocating the second suffixes, repeat the process, continuing until all the c_1 's, c_2 's and c_3 's are lattice.

(c) Proceed as above until all the letters 3.1 are lattice.

It is obvious that any standard right diagram gives rise to a lattice permutation of the n letters 3.1 if we assume that the c_1 's are the symbols appearing in the first row, the c_2 's those in the second row, and so on. E.g. the diagram

$$\begin{array}{c} 134 \\ 25 \end{array} \text{ gives rise to the permutation } c_1 c_2 c_1 c_1 c_2.$$

An exactly similar procedure is applicable to a standard skew diagram, but the resulting permutation will not necessarily be lattice. E.g. the skew diagram

$$\begin{array}{c} 2 \\ 3 \\ 1 \\ 4 \end{array} \text{ gives rise to the permutation } c_3 c_1 c_2 c_4.$$

Altogether, the skew diagram defined by $[\alpha] = [3^2, 2, 1]$ and $[\beta] = [2^2, 1]$

gives rise to $\phi_{\alpha^{\beta}} = 12$ permutations which are associated with the appropriate lattice permutation according to the following table.

Permutation	Lattice Permutation	Irreducible Representation
$c_1 c_2 c_3 c_4$	$c_1 c_2 c_3 c_4$	(1^4)
$c_1 c_4 c_3 c_2$	$c_1 c_2 c_1 c_1$	$(3, 1)$
$c_4 c_1 c_3 c_2$	$c_1 c_1 c_2 c_1$	
$c_4 c_3 c_1 c_2$	$c_1 c_1 c_1 c_2$	(2^2)
$c_3 c_1 c_4 c_2$	$c_1 c_1 c_2 c_2$	
$c_3 c_4 c_1 c_2$	$c_1 c_2 c_1 c_2$	$(2, 1^2)$ twice
$c_1 c_2 c_4 c_3$	$c_1 c_2 c_3 c_1$	
$c_1 c_3 c_4 c_2$	$c_1 c_2 c_3 c_1$	
$c_1 c_3 c_2 c_4$	$c_1 c_2 c_1 c_3$	
$c_1 c_4 c_2 c_3$	$c_1 c_2 c_1 c_3$	
$c_3 c_1 c_2 c_4$	$c_1 c_1 c_2 c_3$	
$c_4 c_1 c_2 c_3$	$c_1 c_1 c_2 c_3$	

We have in fact determined those irreducible representations $[\gamma]$ of S_n whose diagrams can, in the language of the Littlewood and Richardson theorem, be *built on* the diagram $[\beta]$ to yield $[\alpha]$, or can be *built in* $[\alpha] - [\beta]$. The proof given in SG₁ establishes the validity of this determination of the constants $a_{\lambda\beta\gamma}$. In our example we have

$$\phi_{3^2, 2, 1}^{2^2, 1} = f_{1^4} + f_{3, 1} + f_{2^2} + 2f_{2, 1^2}.$$

It should of course be pointed out that the same skew diagram may be defined by different pairs $[\alpha]$ and $[\beta]$. E. g.

$$\phi_{3^2, 2, 1}^{2^2, 1} = \phi_{4^2, 3, 2}^{3^2, 2, 1}.$$

4. There is yet another approach to $\phi_{\alpha^{\beta}}$ which is of interest. Two formulae have long been known^{*} to connect the f 's of the symmetric group S_l with those of S_{l-1} and S_{l+1} :

$$4.1 \quad f_{a_1, a_2, \dots, a_h} = f_{a_1-1, a_2, \dots, a_h} + f_{a_1, a_2-1, \dots, a_h} + \dots + f_{a_1, a_2, \dots, a_{h-1}},$$

and

$$4.2 \quad (l+1)f_{a_1, a_2, \dots, a_h} = f_{a_1+1, a_2, \dots, a_h} + f_{a_1, a_2+1, \dots, a_h} + \dots + f_{a_1, a_2, \dots, a_{h-1}+1},$$

where the additions or deletions contribute only if they do not disturb the inequalities

^{*} [5] reference Young [21], Part III, pp. 261-2.

$$\alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_h.$$

Let us think of these relations as produced by the operation of adding or subtracting a node from the diagram $[\alpha]$. As such, the operation may be iterated to give

$$4.3 \quad f_{\alpha(l)} = \sum_{\beta} \phi_{\alpha}^{\beta} f_{\beta(m)},$$

which is the same as 1.9, and

$$4.4 \quad (l!/m!)f_{\beta(m)} = \sum_{\alpha} \phi_{\alpha}^{\beta} f_{\alpha(l)},$$

where $[\alpha(l)]$ contains $[\beta(m)]$ and $l = m + n$ as before. Iteration of either 4.3 or 4.4 requires that the ϕ 's satisfy the relation

$$4.5 \quad \phi_{\alpha(l)}^{\beta(m)} = \sum_{\mu} \phi_{\alpha(l)}^{\mu(r)} \cdot \phi_{\mu(r)}^{\beta(m)},$$

where $[\alpha(l)]$ contains $[\mu(r)]$ contains $[\beta(m)]$. Formula 4.4 is essentially numerical and has not the more general significance attributed to 4.3. Formula 4.5 is in fact a significant generalization of 4.3.

Young's proof of 4.1 is based on the enumeration of the standard diagrams. Since the last symbol a_l must occupy the end position in its row and column, after removing it, the remainder must still be standard. The inductive argument applied to the explicit form of f is somewhat involved, however, on account of the fact that one passes from the function $\Delta(x_1, x_2, \cdots, x_h)$ to $\Delta(x_1, x_2, \cdots, x_{h-1})$.

On the other hand the proof of 4.2 is much simpler. Noting that the expression for $f_{\beta(m)}$ is the coefficient of $x_1^{\beta_1+h-1} x_2^{\beta_2+h-2} \cdots x_h^{\beta_h}$ in the expansion of

$$(x_1 + x_2 + \cdots + x_h)^m \Delta(x_1, x_2, \cdots, x_h),$$

where Δ is the familiar alternating function, formula 4.2 is merely the coefficient of

$$4.6 \quad x_1^{\beta_1+h-1+1} x_2^{\beta_2+h-2+1} \cdots x_h^{\beta_h+1} (x_1 + x_2 + \cdots + x_{h+1})$$

in the expansion of

$$4.7 \quad (x_1 + x_2 + \cdots + x_{h+1})^{m+1} \Delta(x_1, x_2, \cdots, x_{h+1}).$$

By iteration, formula 4.4 is the coefficient of

$$4.8 \quad x_1^{\beta_1 + h - 1 + n} x_2^{\beta_2 + h - 2 + n} \cdots x_h^{\beta_h + n} x_{h+1}^{n-1} \cdots x_{h+n-1} (x_1 + x_2 + \cdots + x_{h+n})^n$$

in the expansion of

$$4.9 \quad (x_1 + x_2 + \cdots + x_{h+n})^l \Delta(x_1, x_2, \cdots, x_{h+n}).$$

The extra terms which appear in 4.6 have zero coefficient in 4.7. A similar argument is applicable to 4.8 if the factors $(x_1 + x_2 + \cdots + x_{h+n})$ are thought of as multiplied in succession.*

One would like to have an explicit expression for ϕ analogous to that for f . If none of the appropriate terms in 4.8 had vanishing coefficients in 4.9, then ϕ would be a multinomial coefficient. In general, however, this is not the case, so that

$$4.10 \quad \phi_{\alpha(i)}^{\beta(m)} \leq n! / \Pi(\alpha_i - \beta_i)!,$$

and also $\leq n! / \Pi(\alpha_i^* - \beta_i^*)!$,

from 1.11.

It is worth pointing out that a skew diagram need *not* be *connected*, in the sense that any one of its nodes can be reached from any other by a succession of horizontal and vertical steps. Clearly a right diagram is connected. Let us suppose that $[\alpha] - [\beta]$ breaks up into a number of connected skew diagrams containing respectively p, q, r, \cdots nodes, so that $n = p + q + r + \cdots$. Then a little consideration will show that

$$4.11 \quad \phi_{\alpha(i)}^{\beta(m)} = (n! / p! q! r! \cdots) \phi_{(p)} \cdot \phi_{(q)} \cdot \phi_{(r)} \cdots,$$

where the multinomial coefficient comes from the division of the n symbols into sets of p, q, r, \cdots symbols. After such division, the symbols in any set can be taken in their natural order and arranged in the corresponding skew diagram in just ϕ distinct ways.

In concluding this discussion of the generation of the skew representations it should be pointed out that the idea of representing the addition or subtraction of a node from a diagram as an operator goes back to Young (Q. S. A. IV, §§ 6-11) who used the notation E_r and Δ_r for the operation as applied to the r -th row of a diagram. Later (Q. S. A. VI, § 8) he extended the idea, considering the operator $\Delta_i \Delta_j$ applied to a substitutional expression T . In the Littlewood and Richardson Theorem we have the first significant application of the method. One might say that Nakayama's introduction of the notion of a *hook*, to be discussed in Part II, is the second.

* Cf. [2], p. 68.

PART II.

The Murnaghan-Nakayama Recursion Formula.

5. We now apply these ideas to the concept of a *hook* as introduced by Nakayama.⁴ Consider a diagram $[\alpha]$. A *hook* H_r in $[\alpha]$ is defined in § 1 of N_1 to be any diagram

$$5.1 \quad H_r = [n-r, 1^r], \quad (r = 0, 1, \dots, n-1),$$

whose horizontal (vertical) nodes are the *last* $n-r(r)$ nodes of some row (the appropriate column) of $[\alpha]$. We shall say that this hook is *even* or *odd* according as r is even or odd. It will sometimes be convenient to speak of the horizontal part of H_r of length $n-r$ as the *arm* and of the vertical part of length r as the *leg*. As an example consider the diagram $[\alpha] = [4^2, 3, 2, 1]$:

$$5.2 \quad \begin{array}{cccc} & & & \circ \\ & & & \circ \\ & & \circ & \circ \\ & \circ & \circ & \\ & \circ & & \\ & & & \end{array}$$

There is just one hook of length 6, namely $[3, 1^3]$. If this hook is *removed* from $[\alpha]$, the procedure is to think of those nodes which are isolated as being moved to the left and up to produce the diagram $[3, 2, 1^3]$. On the other hand Nakayama remarks in § 5 of N_1 that it is sometimes necessary to think of removing the *equivalent* 'rim' of the diagram, as indicated in 5.2. This latter interpretation is more significant for the operational approach developed in Part I of the present paper. We shall speak of the rim of the diagram as a *skew hook*, to distinguish it from its equivalent *right hook* (5.1). Clearly, the *skew hook* in 5.2 is precisely the skew diagram $[4^2, 3, 2, 1] - [3, 2, 1^3]$.

Nakayama's purpose in introducing the notion of a hook was to study the modular properties of the characters of the symmetric group. We shall postpone the application of our methods to such modular considerations and consider here only the application of the theory to the proof of the Murnaghan-Nakayama recursion formula, as given in § 9 of N_1 :

Let H_1, H_2, \dots be the totality of hooks of length n in the diagram $T = [\alpha]$, then

$$5.3 \quad \chi(T; P) = \sum (-1)^r \chi(T - H_i; P'),$$

where P is any permutation of S_l which contains a cycle C of length n and P' is the permutation on m symbols obtained by removing C from P .

Clearly, the permutation P is contained in the direct product of S_m and S_n and its character in $[\alpha]$ is given by the analogue of 1.9 or 4.3 for characters:

$$5.4 \quad \chi(T; P) = \chi_a(P) = \sum_{\beta} (\chi(C) \text{ in } [\alpha] - [\beta] \cdot \chi_{\beta}(P')).$$

6. We have from 1 that

$$1.10 \quad [\alpha] - [\beta] = \sum_{\gamma} a\lambda_{\beta}\gamma[\gamma].$$

• Murnaghan⁹ proved that

$$6.1 \quad \chi_{\gamma(n)}(C) = (-1)^r \text{ or } 0,$$

according as $[\gamma]$ is an H_r or not. The following theorem gives the requisite information concerning the $a\lambda_{\beta}\gamma$ in 1.10.

THEOREM. If the skew-diagram $[\alpha] - [\beta]$:

- (i) contains interior nodes of $[\alpha]$, then no hook representation occurs;
- (ii) is a complete skew hook, then there is just one hook representation;
- (iii) is an incomplete hook, then the number of even hook representations is equal to the number of odd ones.

If a node of $[\alpha]$ is not on the rim of the diagram it may be said to be an interior node of $[\alpha]$.

If $[\alpha] - [\beta]$ contains an interior node of $[\alpha]$, then it must contain the configuration

$$6.2 \quad \begin{array}{cc} & \circ \\ & \cdot \\ \cdot & \cdot \end{array},$$

where the interior node is ringed. A little consideration will show that no four elements of H_r can be arranged in the pattern 6.2 without violating 1.5, which proves (i). As an illustration we have:

$$6.3 \quad \phi_{4^2, 3, 2, 1}^{2^3, 1^2} = f_{3, 2, 1} + f_{2^3} + f_{2^2, 1^2} = 30.$$

⁹ [3], p. 462.

If $[\alpha] - [\beta]$ is a skew hook, then certainly the equivalent H_r will appear as an irreducible component of $[\alpha] - [\beta]$; since the number of horizontal and vertical steps is just accounted for, the corresponding $a\lambda\beta^\gamma = 1$, and no other hooks can be built in the skew diagram without violating 1.5. This proves (ii). E. g.:

$$6.4 \quad \phi_{4^2, 3, 2, 1}^{3, 2, 1^3} = f_{3, 1^3} + 2f_{3, 2, 1} + f_3^2 + f_2^3 + f_{2^2, 1^2} = 61,$$

where $H_r = [3, 1^3]$ as in 5.2.

Case (iii) is a little more complicated. If one thinks of the removal of a node from a skew hook $[\alpha] - [\beta]$, two possibilities arise: (a) the node in question may be at one end, or (b) it may be an internal node of the rim of $[\alpha]$. Possibility (a) leaves the hook still a hook on one fewer symbols and so is of no interest here. Possibility (b) relaxes two conditions, since the removal must be at a corner of the skew diagram. Two hooks H_{r_1} and H_{r_2} may be built in the resulting skew diagram, which we may call an *incomplete* skew hook, one obtained by removing a node from the arm of H_r , the other by removing a node from its leg. As an illustration of the process we have:

$$6.5 \quad \phi_{4^2, 3, 2, 1}^{4, 3, 1} = f_{4, 2} + f_{4, 1^2} + 3f_{3, 2, 1} + f_3^2 + f_{3, 1^3} + f_2^3 + f_{2^2, 1^2} = 96,$$

where $H_r = [4, 1^3]$, $H_{r_1} = [4, 1^2]$ and $H_{r_2} = [3, 1^3]$. One of H_{r_1} and H_{r_2} is consequently *even* while the other is *odd*.

If now a second node be removed from this incomplete skew hook, it may be next to the first node or it may be another internal node of the rim of $[\alpha]$. In the first case the only possible hooks which can be built in $[\alpha] - [\beta]$ are obtained by removing a node from the arms of *both* H_{r_1} and H_{r_2} or from their legs; one of the resulting hooks will be even while the other one will be odd. The removal of a second internal node again relaxes two conditions and makes it possible to build hooks obtained by removing a node from *each* of the arms and legs of H_{r_1} and H_{r_2} ; of the resulting four hooks, two will be even and two odd. E. g. we have:

$$6.6 \quad \phi_{4^2, 3, 2, 1}^{3^2, 2} = f_{4, 2} + f_{4, 1^2} + f_3^2 + 3f_{3, 2, 1} + 2f_{3, 1^3} + f_2^3 + 2f_{2^2, 1^2} + f_{2, 1^4} = 120,$$

where $[4, 1^2]$, $[3, 1^3]$ and $[3, 1^3]$, $[2, 1^4]$ are obtained as above described from $[4, 1^3]$ and $[3, 1^4]$. Further repetition of the argument proves (iii).

7. The Murnaghan-Nakayama recursion formula now follows immediately from 6.1, since

$$7.1 \quad \chi(C) \text{ in } [\alpha] - [\beta] = 0$$

in cases (i) and (iii), while

$$7.2 \quad \chi(C) \text{ in } [\alpha] - [\beta] = (-1)^r$$

in case (ii).

Taking $n = p$, it follows from the modular theory that

$$7.3 \quad f_{\gamma(p)} \equiv (-1)^r \text{ or } 0, \text{ mod } p,$$

according as $[\gamma]$ is an H_r or not. Formula 7.3 should be compared with 6.1. As before

$$7.4 \quad \phi_{a(i)}^{\beta(m)} = \sum_{\gamma} a\lambda_{\beta} \gamma f_{\gamma(p)} \equiv (-1)^r \text{ or } 0, \text{ mod } p,$$

according as $[\alpha] - [\beta]$ is a skew hook equivalent to H_r or not. The analogy between the f 's and the ϕ 's is striking throughout.

THE UNIVERSITY OF TORONTO.

REFERENCES.

-
- [1] Stig Comét, "Une propriété des déterminantes et son application au calcul des caractères des groupes symétriques," *Kungl. Fysiografiska Sällskapet i Lund Förhandlingar* 14, no. 7, pp. 84-94 (1945).
 - [2] D. E. Littlewood, *The Theory of Group Characters*, Oxford, 1940.
 - [3] F. D. Murnaghan, "On the representations of the symmetric group," *American Journal of Mathematics*, vol. 59 (1937), pp. 437-488.
 - [4] T. Nakayama, "On some modular properties of irreducible representations of a symmetric group," Part I, *Japanese Journal of Mathematics*, vol. 17 (1940), pp. 89-108. Part II, same volume, pp. 411-423.
 - [5] G. de B. Robinson, "On the representations of the symmetric group," *American Journal of Mathematics*, vol. 60 (1938), pp. 745-760.
 - [6] R. M. Thrall, "Young's semi-normal representation of the symmetric group," *Duke Mathematical Journal*, vol. 8 (1941), pp. 611-624.
 - [7] H. Weyl, *The Classical Groups*, Princeton, 1939.

POLYNOMIAL MATRICES IN ONE VARIABLE, DIFFERENTIAL EQUATIONS AND MODULE THEORY.*

By ERNST SNAPPER.

Introduction and Summary. Let $P[x]$ be the ring of polynomials in one variable x with coefficients in the commutative field P . Let A be an $m \times s$ matrix of rank r whose elements belong to $P[x]$. Let Δ_i be the highest common factor of the $i \times i$ subdeterminants of A for $i = 1, \dots, r$. The purpose of Part I of this paper is to develop the theory of the norm $\Delta = \Delta_r$ and the elementary divisor $\epsilon = \Delta_r : \Delta_{r-1}$ intrinsically in terms of the column space M of A (the column space of a matrix is the vector space generated by its columns). The purpose of Part II is to extend this theory to the case where M is any module with a Noetherian ring (a ring in which every ideal has a finite basis) as operator domain.

In Part I it is shown how Δ and ϵ can be defined, by the use of composition sequences and other special sequences, intrinsically in terms of M , thus making it obvious that two matrices with the same column spaces have the same Δ and ϵ . These definitions are then used to investigate the properties of Δ and ϵ (see 1.1) and to develop the theory of a system of linear algebraic equations with coefficients in $P[x]$ (see 1.2 and 1.3) and of a system of homogeneous linear differential equations with constant coefficients (see 1.4-1.6). This treatment of the matrix A differs from the classical theory in the following respects:

1) In the classical theory Δ and ϵ are defined in terms of determinants of A and not intrinsically in terms of M (see [1], p. 91 or [2], p. 27; square brackets refer to the references).

2) In the classical theory of the above mentioned differential equations, all the determinantal factors $\Delta_r, \Delta_{r-1}, \dots, \Delta_1$ or elementary divisors $\Delta_r : \Delta_{r-1}, \Delta_{r-1} : \Delta_{r-2}, \dots, \Delta_1$ occur (see [1], chapter 5) while the matrix of the system is assumed to be square and non-singular. In 1.4-1.6 this theory is derived by the use of only Δ and ϵ while composition sequences of M occur instead of determinantal factors or elementary divisors of order less than r . Furthermore, in 1.4 the notion of "trivial solution" is introduced which enables us to develop the theory for a system with an arbitrary $m \times s$ matrix instead of

* Received October 4, 1945.

only for a system with a square non-singular matrix. In these three sections it is assumed that P is the field of complex numbers, not from mathematical necessity but in the hope that these sections will also be read by non-algebraists who are interested in differential equations.

In Part II it is shown how, by the use of the module theory developed by P. M. Grundy (see [5]) and the author (see [6]), the theory of Part I can be extended to the case where M is any module with a Noetherian ring \mathfrak{s} as operator domain. Sections 2.1-2.4 review those parts of this module theory, necessary for the extensions. Section 2.5 shows how the notions of Δ and ϵ can be extended. Sections 2.6-2.8 deal with the theory of a system of algebraic linear equations

$$(1) \quad \sum_{j=1}^s \alpha_{ij} x_j = \gamma_i, \quad i = 1, \dots, m$$

where α_{ij} and γ_i are given elements of \mathfrak{s} . The "criterion of solvability" which says when (1) can be solved by elements $x_j \in \mathfrak{s}$ is based on the "criterion of lengths" (see 2.6) which is the module analogue of the "Langensatz" of ideal theory (see [7], p. 35). It is shown in 2.7 how the "criterion of lengths" becomes the classical theorem of solvability of the system (1) in the cases where \mathfrak{s} is specialized to be a field (see [2], p. 70) or a principal ideal ring or $P[x]$ (see [3], p. 60) or the integers of an algebraic number field (see [4], p. 340).

A paper to be published in this journal under the name "Polynomial Matrices in several Variables," referred to here as MSV, will treat the special case where $\mathfrak{s} = P[x_1, \dots, x_n]$, that is a polynomial ring in several variables. Frequent remarks throughout the present paper point out connections between Part I, Part II, MSV, [5] and [6]; they are not a logical part of the development of the theory.

The author is not able to extend the notions of the determinantal factors $\Delta_{r-1}, \dots, \Delta_1$ and elementary divisors $\Delta_{r-1} : \Delta_{r-2}, \dots, \Delta_1$ of order less than r in an analogous way beyond the case where $\mathfrak{s} = P[x]$. Such an extension would be useful only if it supplied a criterion by which one could tell when two modules are operator isomorphic; such a criterion is indeed supplied by the elementary divisors when $\mathfrak{s} = P[x]$.

Part I. Polynomial Matrices in One Variable.

1.1. Closure, associated primes, elementary divisor and norm of a module. Let $P[x]$ be the polynomial ring consisting of the polynomials in one variable x whose coefficients are elements of the arbitrary, commutative,

field P . Elements of $P[x]$ are called *scalars* and elements of P *constants*. Let V be the m -dimensional vector space V over $P[x]$, consisting of the column vectors whose m components are elements of the scalar domain $P[x]$. The vectors of V can be added, subtracted and multiplied by scalars in the usual way. A *module* is a subset of V which is closed under vector subtraction and scalar multiplication. If, without further explanation, a capital Roman letter, a lower case Roman letter or a lower case Greek letter is used, the letter always denotes, respectively, a module, a vector or a scalar. The ideal of $P[x]$, generated by α , is denoted by (α) and the module, generated by u, v, \dots, w , is denoted by (u, v, \dots, w) . In the same way (M, u, v, \dots, w) denotes the module which is generated by the vectors of M together with u, v, \dots, w ; i. e., $h \in (M, u, v, \dots, w)$ if $h = m + \alpha u + \beta v + \dots + \gamma w$ where $m \in M$. We shall never distinguish between two scalars ϕ_1 and ϕ_2 if $\phi_1 = \alpha \phi_2$ where α is a constant. Hence, we shall speak about "the" generator of an ideal of $P[x]$ and shall never worry about the fact that such a generator is determined only up to a multiplicative constant. The symbol $M : \alpha$ denotes the module consisting of all vectors v such that $\alpha v \in M$. Clearly, $M : \alpha \supseteq M$. The symbol $M : N$ denotes the ideal consisting of all scalars α such that $\alpha n \in M$ for all $n \in N$. Throughout this paper, the symbols \subset and \supset will be used exclusively for proper inclusion.

DEFINITION 1.11. *An irreducible scalar $\pi \neq 0$ is a non-zero associated prime of M if and only if $M : \pi \supset M$.*

Remark 1.11. The non-zero associated primes of Definition 1.11 are identical with those defined by the Noether decomposition of M (see Remark 2.23). For the zero associated prime of a module, see Remark 2.32.

DEFINITION 1.12. *The closure $\text{Cl}(M)$ is the module which consists of all vectors v such that $\alpha v \in M$ for some non-zero scalar. The elementary divisor ϵ of M is the generator of the ideal $M : \text{Cl}(M)$.*

Clearly $\epsilon \neq 0$, since if $\text{Cl}(M) = (v_1, \dots, v_s)$ there exist non-zero scalars $\alpha_1, \dots, \alpha_s$ such that $\alpha_i v_i \in M$ for $i = 1, \dots, s$. The product $\alpha_1 \dots \alpha_s$ is then a non-zero element of $M : \text{Cl}(M)$, hence $\epsilon \neq 0$. It follows immediately that $M : \epsilon = \text{Cl}(M)$. Furthermore always $M \subseteq \text{Cl}(M)$. We say that M is *closed* if $M = \text{Cl}(M)$ and that M is *dense* if $\text{Cl}(M) = V$.

Remark 1.12. If A is a matrix whose columns generate M , then ϵ of M is equal to the usual elementary divisor of highest rank of A (see [3], Theorem II).

The following theorem shows that the number of non-zero associated primes of M is finite.

THEOREM 1.11. *If M is not closed, the non-zero associated primes are the irreducible factors of ϵ . Furthermore, the following three statements are equivalent: (1) M is closed; (2) ϵ is a non-zero constant; (3) M has no non-zero associated primes.*

Proof. If π is a non-zero associated prime of M , a $v \notin M$ exists such that $\pi v \in M$ and $\epsilon v \in M$. If π were not a factor of ϵ , we could find scalars α and β such that $(\alpha\pi + \beta\epsilon)v = v$ which would imply $v \in M$. Hence, every non-zero associated prime is a factor of ϵ . Conversely, let π be an irreducible factor of ϵ ; hence $\epsilon = \pi^o\phi$ and, if $v \in \text{Cl}(M)$, $\pi^o\phi v \in M$. Now, if $M : \pi = M$, then we could conclude from $\pi^o\phi v \in M$ that $\phi v \in M$. Since ϵ is the highest common factor of all scalars which transform all vectors of $\text{Cl}(M)$ into vectors of M , this would imply that $\phi = \epsilon$ against hypothesis. Consequently, $M : \pi \supset M$ and π is a non-zero associated prime of M . For the second part of the theorem, suppose that M is closed. Then $M : \text{Cl}(M) = M : M = P[x]$ and hence ϵ is a non-zero constant. Conversely, if ϵ is a non-zero constant and $v \in \text{Cl}(M)$, then from $\epsilon v \in M$ follows $v \in M$ and hence $\text{Cl}(M) = M$. Consequently, (1) and (2) are equivalent. We now show that (1) and (3) are equivalent. If M is closed, $M : \alpha = M$ for any non-zero scalar α and hence (3) is satisfied. Conversely, if (3) is satisfied, M must be closed for, if $M \subset \text{Cl}(M)$, then $M : \text{Cl}(M) \subset P[x]$ and hence $\epsilon \notin P$. Every irreducible factor of ϵ is then a non-zero associated prime of M according to the first part of the theorem, against the hypothesis that (3) holds. This proves Theorem 1.11.

Let $\epsilon = \prod_{j=1}^h \pi_j^{\rho_j}$. The meaning of the irreducible factors π_j has been explained. The following theorem explains the meaning of the multiplicity ρ_j .

THEOREM 1.12. *The sequence (1) $M \subset M : \pi_j \subset \dots \subset M : \pi_j^{\rho_j}$ consists of $\rho_j + 1$ different terms. For $h \geq 1$, $M : \pi_j^{\rho_j+h} = M : \pi_j^{\rho_j}$.*

Proof. In order to show that the $\rho_j + 1$ terms of (1) are different, we observe that $M : \pi_j^{k+1} = (M : \pi_j^k) : \pi_j$ and hence that all we have to show is $M : \pi_j^{\rho_j-1} \subset M : \pi_j^{\rho_j}$. Since $\epsilon = \pi_j^{\rho_j}\phi$, we can find a $v \in \text{Cl}(M)$ such that $\pi_j^{\rho_j-1}\phi v \notin M$ while of course $\pi_j^{\rho_j}\phi v \in M$, for, otherwise, $\epsilon = \pi_j^{\rho_j-1}\phi$. Hence, $\phi v \in M : \pi_j^{\rho_j}$ but $\phi v \notin M : \pi_j^{\rho_j-1}$ which proves $M : \pi_j^{\rho_j-1} \subset M : \pi_j^{\rho_j}$. Now, let $v \in M : \pi_j^{\rho_j+h}$ where $h \geq 1$. Then, $\pi_j^{\rho_j+h}v \in M$ and $\pi_j^{\rho_j}\phi v \in M$. Since π_j and ϕ are relatively prime we can find scalars α and β such that $(\alpha\pi_j^h + \beta\phi)v = v$ which implies $\pi_j^{\rho_j}v = \alpha\pi_j^{\rho_j+h}v + \beta\pi_j^{\rho_j}\phi v \in M$ and hence $v \in M : \pi_j^{\rho_j}$.

We now turn our attention to the norm Δ of M by refining the sequence (1) to a composition sequence. Here, a sequence $M_0 = M \subset M_1 \subset \cdots \subset M_l$ is called a *composition sequence from M to M_l of length l* if no module N exists such $M_i \subset N \subset M_{i+1}$ for some $0 \leq i \leq l-1$. Clearly, the factor modules $0 \subset M_1/M \subset \cdots \subset M_l/M$ form an ordinary composition sequence of length l of M_l/M (see [9], p. 150). Conversely, if $0 \subset C_1 \subset \cdots \subset M_l/M$ is a composition sequence of length l of M_l/M and if M_i consists of all the vectors of the coset C_i , then $M \subset M_1 \subset \cdots \subset M_l$ is a composition sequence from M to M_l of length l .

THEOREM 1.13. *If $M \subset N \subseteq \text{Cl}(M)$ and if $M \subset M' \subset \cdots \subset N$ is any sequence from M to N , this sequence can be refined to a composition sequence from M to N . The length l of such a composition sequence depends only on M and N and, if $M_i \subset M_{i+1}$ are two consecutive members of the sequence, the factor module M_{i+1}/M_i has finite rank r_i with respect to P . Consequently, N/M has finite rank $r = \sum_{i=1}^l r_i$ with respect to P .*

Proof. For "rank with respect to P " we shall write P -rank. Theorem 1.13 follows from the theory of composition sequences of $\text{Cl}(M)/M$ as soon as we have proved the existence of one composition sequence from M to $\text{Cl}(M)$, of which two consecutive members give rise to a factor module which has finite P -rank. Let $\epsilon = \Pi_{j=1}^h \pi_j^{\rho_j}$ and consider the sequence $M \subset M : \pi_1^{\rho_1} \subset M : \pi_1^{\rho_1} \pi_2^{\rho_2} \subset \cdots \subset M : \epsilon = \text{Cl}(M)$. If $M_i = M : \pi_1^{\rho_1} \cdots \pi_i^{\rho_i}$, then $M_{i+1} = M_i : \pi_{i+1}^{\rho_{i+1}}$ for $0 \leq i \leq h-1$ and the sequence $M_i \subset M_{i+1}$, formed by two consecutive members, can be refined to $M_i \subset M_i : \pi_{i+1} \subset M_i : \pi_{i+1}^2 \subset \cdots \subset M_i : \pi_{i+1}^{\rho_{i+1}} = M_{i+1}$. We show that we can draw a composition sequence from $M_i : \pi_{i+1}^k$ to $M_i : \pi_{i+1}^{k+1}$ for $0 \leq k \leq \rho_{i+1}-1$. Suppose that $v \in M_i : \pi_{i+1}^{k+1}$ and $v \notin M_i : \pi_{i+1}^k$ and let $L = (M_i : \pi_{i+1}^k, v)$. Since $\alpha v \in M_i : \pi_{i+1}^k$ if and only if $\alpha \in (\pi_{i+1})$, the factor module $L/M_i : \pi_{i+1}^k$ is operator isomorphic with $P[x]/(\pi_{i+1})$ with respect to $P[x]$ as operator domain. Since (π_{i+1}) is a maximal prime ideal of $P[x]$, $P[x]/(\pi_{i+1})$ has no proper non-zero sub-ideal and has finite P -rank ∂_{i+1} if ∂_{i+1} is the degree of the polynomial π_{i+1} . Hence, no proper sub-module can be included between M and L and ∂_{i+1} is the P -rank of L/M . If $L = M_i : \pi_{i+1}^{k+1}$, the proof is complete; otherwise we choose a vector $w \in M_i : \pi_{i+1}^{k+1}$ and $w \notin L$ and repeat the process. Because the ascending chain condition holds in V , we must reach $M_i : \pi_{i+1}^{k+1}$ in a finite number of steps and Theorem 1.13 is proved.

Since $M : \pi_j^{\rho_j} \subseteq \text{Cl}(M)$, we can draw composition sequences from M to $M : \pi_j^{\rho_j}$ according to Theorem 1.13. In the remainder of Part I, l_j will

always denote the length of a composition sequence from M to $M : \pi_j^{\rho_j}$ and ∂_j the degree of the polynomial π_j for $1 \leq j \leq h$.

THEOREM 1.14. *The P -rank of $M : \pi_j^{\rho_j}/M$ is $l_j \partial_j$ and of $\text{Cl}(M)/M$ is $\partial = \sum_{j=1}^h l_j \partial_j$.*

Proof. According to the notation of the previous theorem, two consecutive members $M_i \subset M_{i+1}$ of the sequence $M \subset M_1 \subset \cdots \subset M_h = \text{Cl}(M)$ were refined to a composition sequence of length l'_{i+1} , and $l'_{i+1} \partial_{i+1}$ was the P -rank of M_{i+1}/M_i . Hence, all we have to show is that $l'_{i+1} = l_{i+1}$. Let $M'_j \subset M'_{j+1}$ be two consecutive members of a composition sequence from M to $M : \pi_{i+1}^{\rho_{i+1}}$ and let $\pi_1^{\rho_1} \cdots \pi_i^{\rho_i} = \phi_i$. Then, $M'_{j+1} = (M'_j, v_j)$ and $\alpha v_j \in M'_j$ if and only if $\alpha \in (\pi_{i+1})$. Since ϕ_i and π_{i+1} are relatively prime, we can find scalars β and γ such that $(\beta \phi_i + \gamma \pi_{i+1}) v_j = v_j$. Hence $M'_{j+1} = (M'_j, \phi_i w_j)$ where $w_j = \beta v_j$. The composition sequence from M to $M : \pi_{i+1}^{\rho_{i+1}}$ can hence be written as $M \subset (M, \phi_i w_1) \subset (M, \phi_i w_1, \phi_i w_2) \subset \cdots \subset (M, \phi_i w_1, \cdots, \phi_i w_l) = M : \pi_{i+1}^{\rho_{i+1}}$. It can then be proved without difficulty that $M'_i \subset (M_i, w_1) \subset (M_i, w_1, w_2) \subset \cdots \subset (M_i, w_1, \cdots, w_l) = M_i : \pi_{i+1}^{\rho_{i+1}}$ is a composition sequence from M_i to $M_i : \pi_{i+1}^{\rho_{i+1}}$ and hence that $l'_{i+1} = l_{i+1}$.

DEFINITION 1.13. *The norm Δ of a module M is defined as $\Delta = \Pi_{j=1}^h \pi_j^{l_j}$.*

Since ρ_j is the length of an ordinary sequence (without repetition) from M to $M : \pi_j^{\rho_j}$ and l_j is the length of a composition sequence, $l_j \geq \rho_j$ and hence ϵ is always a divisor of Δ . According to Theorem 1.14, the degree of Δ is ∂ , that is the P -rank of $\text{Cl}(M)/M$.

Remark 1.13. If A is a matrix whose columns generate M , Δ of M is equal to the determinantal factor of highest rank of A as follows easily from [3], pp. 60 and 61. The difference in character between ϵ and Δ is now clear. The multiplicities of Δ are "lengths" and hence are closely related to P -ranks, which makes Δ fundamental for systems of algebraic equations. The multiplicities of ϵ , however, tell when certain sequences break off, which makes ϵ fundamental for systems of differential equations.

Remark 1.14. The composition sequences from M to $M : \pi_j^{\rho_j}$ correspond to the primary composition sequences of the general module theory (see 2.4 and 2.5) which, in their turn, correspond to the primary composition sequences of ideal theory (see [10], Section 23).

1.2. Rank and closure. Let M be a module, $\text{Cl}(M)$ its closure, and r its rank; i.e., r is the maximal number of vectors of M which are linearly

independent with respect to $P[x]$. Hence, rank, without further explanation, always means $P[x]$ -rank. The following theorem connects the notions of closure and rank.

THEOREM 1.21. *$\text{Cl}(M)$ is closed, has the same rank r as M , and contains every module whose rank is r and which contains M . Furthermore, $\text{Cl}(M_1 \wedge M_2) = \text{Cl}(M_1) \wedge \text{Cl}(M_2)$ and, if $M_1 \subseteq M_2$, then $\text{Cl}(M_1) \subseteq \text{Cl}(M_2)$.*

Proof. The first half of the theorem is proved in [3], p. 56 and the second half in [6], p. 259.

It follows from Theorem 1.21 that $\text{Cl}(M)$ could have been defined as the only closed module of rank r which contains M . Consequently, if $M_1 \subseteq M_2$ and M_1 and M_2 have the same rank, then $\text{Cl}(M_1) = \text{Cl}(M_2)$. Finally, M is dense if and only if $r = m$.

Remark 1.21. Theorem 1.21 is used in 1.3 for the theory of systems of linear algebraic equations. In the case of arbitrary modules the notion of rank is replaced by the more general notion of length of a primary sequence (see 2.7 and 2.8).

DEFINITION 1.21. *If A is an $m \times s$ matrix whose columns are vectors of V , then the closure, rank, elementary divisor, norm and associated primes of A are defined as those of the column space of A , i. e., of the module M generated by the columns of A .*

1.3. Systems of linear equations.

THEOREM 1.31. *Let $M_1 \subseteq M_2$ be two modules. Then, $M_1 = M_2$ if and only if M_1 and M_2 have the same rank and norm.*

Proof. It is obvious that if $M_1 = M_2$ the norms and ranks of M_1 and M_2 are the same. If $M_1 \subseteq M_2$ and the ranks are equal, then $\text{Cl}(M_1) = \text{Cl}(M_2)$. Let $M_2 \subset M_{21} \subset \cdots \subset M_{2l} = \text{Cl}(M_2)$ be a composition sequence from M_2 to $\text{Cl}(M_2)$ of length l and let r_i be the P -rank of the factor module $M_{2,i+1}/M_{2,i}$. Then, $\partial = \sum_{i=1}^l r_i$ is the P -rank of $\text{Cl}(M_2)/M_2$ and hence is equal to the degree of the norm of M_2 . If $M_1 \subset M_2$, the sequence $M_1 \subset M_2 \subset M_{21} \subset \cdots \subset M_{2l} = \text{Cl}(M_1)$ would prove that the P -rank of $\text{Cl}(M_1)/M_1$ is $\partial + r_0$ where r_0 is the P -rank of M_2/M_1 . Hence, if also the norms of M_1 and M_2 are the same, $r_0 = 0$ and $M_1 = M_2$.

Theorem 1.31 contains the criterion, given in Theorem 1.32, for the

solvability of a system of linear equations (Theorem 1.32 is the analogue of the theorem on p. 340 in [4]).

THEOREM 1.32. *Let $\sum_{j=1}^s \alpha_{ij} z_j = \gamma_i$, where $i = 1, \dots, m$ and α_{ij} and γ_i are polynomials of $P[x]$, be a system of m linear equations for the s unknowns z_j . Then, this system has a solution $z_j \in P[x]$ ($j = 1, \dots, s$) if and only if the $m \times s$ matrix $A = (\alpha_{ij})$ and the augmented $m \times (s+1)$ matrix $B = (\alpha_{ij}, \gamma_i)$ have the same rank and norm.*

Proof. Let M be the column space of A and g the vector whose components are $\gamma_1, \dots, \gamma_m$. Then the system has a solution if and only if $g \in M$, i. e., if and only if $M = L$, where $L = (M, g)$. Since $M \subseteq L$, Theorem 1.32 is an immediate corollary of Theorem 1.31 and Definition 1.21.

1.4. Systems of differential equations. We shall investigate the exponential solutions of a system S of linear, homogeneous, differential equations with constant coefficients.

Let W be the m -dimensional vector space consisting of the column vectors whose components are elements of the polynomial ring $P[t]$, where t is a new variable. For reasons given in the introduction, we assume in 1.4, 1.5 and 1.6 that P is the field of complex numbers. An exponential vector $w \exp(\xi t)$ is the product of a vector $w \in W$ and an exponential $\exp(\xi t) = e^{\xi t}$, where $\xi \in P$ and where e is the basis of the natural logarithms. The derivative with respect to t , $Dw \exp(\xi t)$, is defined as in analysis where it is proved that $Dw \exp(\xi t) = \exp(\xi t) (D + \xi)w$ (see [1], Chapter V). If it is important to indicate the independent variable, we will write $A(x)$ instead of A and $w(t)$ instead of $w \in W$. For $v \in V$ we may write $v(x)$, and $v(t)$ then arises from $v(x)$ by the substitution $x = t$. In the same way, the operator vector $v(D)$ is the vector whose components are differential operators which arise from $v(x)$ or $v(t)$ by the substitutions $x = D$ and $t = D$ respectively. We assume in 1.4, 1.5 and 1.6 that A is a fixed $m \times s$ matrix of rank r whose columns are vectors of V . The operator matrix $A(D)$ arises from A by the substitution $x = D$. The system S whose exponential solutions we want to investigate is given by $A(D)y = 0$. Hence we want to discuss the exponential vectors $y = w \exp(\xi t)$ such that $A(D)w \exp(\xi t) = 0$.

An exponential vector $w \exp(\xi t)$ is said to be of degree κ if κ is the degree of w , where the degree of a vector is the highest degree of the components of the vector. Hence, a vector $w(t)$ is of degree κ if $w = \sum_{j=0}^{\kappa} c_j t^j$ where the c_j 's are constant vectors (i. e., the components of c_j are elements of P) and where $c_{\kappa} \neq 0$. The letter c (or c_j etc.), used without further

explanations, will always denote a constant vector or a constant, depending on the connotation.

LEMMA 1.41. *Every exponential vector $w \exp(\xi t)$ of degree κ can be written as $u(D)t^\kappa \exp(\xi t)$, where $u(x) \in V$. Then, in the expansion $u(x) = \sum_j c_j x^j$, the $c_0, c_1, \dots, c_\kappa$ are uniquely determined by $w \exp(\xi t)$ and c_j for $j > \kappa$ is arbitrary and hence may be made zero; furthermore, always $u(\xi) \neq 0$. Conversely, an expression $u(D)t^\kappa \exp(\xi t)$ always represents an exponential vector of degree $\leq \kappa$. The degree is κ if and only if $u(\xi) \neq 0$.*

Proof. The symbol $u(\xi)$ denotes of course the constant vector which arises from $u(x)$ by the substitution $x = \xi$. In the same way let $u^{(j)}(\xi)$ denote the vector which arises from the j -th derivative of $u(x)$ with respect to x by the substitution $x = \xi$. The Taylor expansion then gives $u(D)t^\kappa \exp(\xi t) = \exp(\xi t)u(D + \xi)t^\kappa = \exp(\xi t)\sum_j (1/j!)u^{(j)}(\xi)D^j t^\kappa = \exp(\xi t)\sum_j \binom{\kappa}{j} u^{(j)}(\xi)t^{\kappa-j}$, where $\binom{\kappa}{j}$ is the binomial coefficient $\kappa!/j!(\kappa-j)!$. The lemma follows immediately by comparing this Taylor expansion with the expansion of $w \exp(\xi t) = \exp(\xi t)\sum_{j=0}^\infty d_j t^j$ where the d_j are constant vectors and $d_\kappa \neq 0$.

We now return to the system S . The null space N of $A(x)$ is defined as the module of V which consists of the vectors $n(x)$ such that $A(x)n(x) = 0$. If $n \in N$, clearly $A(D)n(D)t^\kappa \exp(\xi t) = 0$ for any κ and ξ . This leads to the following definition.

DEFINITION 1.41. *An exponential solution $w \exp(\xi t)$ of S is called trivial if there exists an $n \in N$ such that $w \exp(\xi t) = n(D)t^\kappa \exp(\xi t)$. Otherwise, the solution is called non-trivial.*

Observe that, if $w \exp(\xi t)$ is a trivial solution of S of degree κ , $n \in N$ can always be chosen in such a way that $w \exp(\xi t) = n(D)t^\kappa \exp(\xi t)$, for there exists some $n \in N$ such that $w \exp(\xi t) = n(D)t^\sigma \exp(\xi t)$ for $\sigma \geq \kappa$. If $\sigma > \kappa$, then $n(\xi) = 0$ according to Lemma 1.41. Consequently, then $n(x) = (x - \xi)n^*(x)$ where $n^*(x)$ is again a vector of N since null spaces are clearly always closed modules. Consequently, $n(D)t^\sigma \exp(\xi t) = n^*(D)(D - \xi)t^\sigma \exp(\xi t) = \sigma n^*(D)t^{\sigma-1} \exp(\xi t)$ where $\sigma n^*(x) \in N$. If $\sigma - 1 > \kappa$ we can repeat this process until we find a vector $n' \in N$ such that $w \exp(\xi t) = n'(D)t^\kappa \exp(\xi t)$.

The following theorem establishes the relationship between the exponential solutions of S and the theory of the previous sections. The column space M of A , its invariants $\text{Cl}(M)$, Δ , ϵ , r , the non-zero associated primes $\pi_j = x - \xi_j$ ($j = 1, \dots, h$), and the multiplicities ρ_j of $\epsilon = \prod_{j=1}^h (x - \xi_j)^{\rho_j}$ and l_j of

$\Delta = \Pi_{j=1}^h (x - \xi_j)^{l_j}$ are all fixed. Since $\partial_j = 1$ (see Theorem 1.14), l_j is the P -rank of $M : (x - \xi_j)^{\rho_j}/M$ and $\partial = \Sigma_{j=1}^h l_j$ is the P -rank of $\text{Cl}(M)/M$.

THEOREM 1.41. *The exponential vector $w \exp(\xi t) = u(D)t^{\kappa-1} \exp(\xi t)$ is an exponential solution of S , where $\kappa \geq 1$, if and only if $A(x)u(x) = (x - \xi)^\kappa v(x)$. The solution is non-trivial if and only if $v(x) \notin M$. This implies that, if $w \exp(\xi t)$ is a non-trivial solution of S , $x - \xi$ is a non-zero associated prime of M , i. e., ξ is then a root of ϵ .*

Proof. $A(D)u(D)t^{\kappa-1} \exp(\xi t) = \exp(\xi t) \Sigma_{j=0}^{\kappa-1} \binom{\kappa-1}{j} (Au)^{(j)}(\xi) t^{\kappa-1-j}$, where $(Au)^{(j)}(\xi)$ denotes the result of taking the j -th derivative of the vector $A(x)u(x)$ and then substituting $x = \xi$. Hence, $u(D)t^{\kappa-1} \exp(\xi t)$ is a solution of S if and only if $(Au)^{(j)}(\xi) = 0$ for $j = 0, \dots, \kappa - 1$, i. e., if and only if $A(x)u(x) = (x - \xi)^\kappa v(x)$. If $v(x) \in M$ the solution would be trivial, for then $v(x) = Af(x)$ and hence $u = (x - \xi)^\kappa f + n$, where $n \in N$. This would imply $u(D)t^{\kappa-1} \exp(\xi t) = f(D)(D - \xi)^\kappa t^{\kappa-1} \exp(\xi t) + n(D)t^{\kappa-1} \exp(\xi t) = n(D)t^{\kappa-1} \exp(\xi t)$ which proves the triviality. Conversely, if the solution is trivial, there exists an $n \in N$ such that $u(D)t^{\kappa-1} \exp(\xi t) = n(D)t^{\kappa-1} \exp(\xi t)$. From Lemma 1.41 we conclude that then $u(x)$ and $n(x)$ are the same except for terms of degree $\geq \kappa$ and hence $u(x) = n(x) + (x - \xi)^\kappa g(x)$. This implies that $A(x)u(x) = (x - \xi)^\kappa A(x)g(x) = (x - \xi)^\kappa v(x)$ and consequently that $A(x)g(x) = v(x)$, i. e., that $v \in M$. Finally, if $w \exp(\xi t)$ is a non-trivial solution of S , $v(x) \in M : (x - \xi)^\kappa$ and $v \notin M$ from which it follows that $M : (x - \xi)^\kappa \supset M$ and hence that $M : (x - \xi) \supset M$. According to Definition 1.11, this means that $x - \xi$ is then a non-zero associated prime of M .

Hence, for the discussion of the non-trivial solutions of S we have to consider only the roots ξ_1, \dots, ξ_h of ϵ . Consider the sequence $M \subset M : (x - \xi_j) \subset \dots \subset M : (x - \xi_j)^{\rho_j}$ of 1.1 for some arbitrarily chosen but fixed root ξ_j of ϵ and let l_κ be the P -rank of the factor module $M : (x - \xi_j)^\kappa / M : (x - \xi_j)^{\kappa-1}$ for $1 \leq \kappa \leq \rho_j$. The P -rank l_κ depends of course on the choice of ξ_j but the index j will always be omitted except for the invariants which were considered in the previous section. Then, $\Sigma_{\kappa=1}^{\rho_j} l_\kappa = l_j$, and, for each $1 \leq \kappa \leq \rho_j$, we can choose l_κ fixed vectors $v_{\kappa 1}, \dots, v_{\kappa l_\kappa}$ of $M : (x - \xi_j)^\kappa$ which are P -linearly independent mod $M : (x - \xi_j)^{\kappa-1}$. Naturally, P -linearly independent or dependent stands for linearly independent or dependent with respect to P ; consequently, if $\Sigma_{i=1}^{l_\kappa} c_i v_{\kappa i} \in M : (x - \xi_j)^{\kappa-1}$, all the l_κ constants c_i are zero. For each of the l_j vectors $v_{\kappa i}$ we choose some fixed vector $u_{\kappa i}$ such that $Au_{\kappa i} = (x - \xi)^\kappa v_{\kappa i}$, which can be done since $v_{\kappa i} \in M : (x - \xi)^\kappa$. Finally, with each $u_{\kappa i}$ we associate the vector $w_{\kappa i} \in W$ which is uniquely determined by

$w_{\kappa i} \exp(\xi_j t) = u_{\kappa i}(D)t^{\kappa-1} \exp(\xi_j t)$. The following theorem reveals the structure of the non-trivial exponential solutions $w \exp(\xi_j t)$ of S .

THEOREM 1.42. *Let the vectors $v_{\kappa i}$, $u_{\kappa i}$, $w_{\kappa i}$ for $1 \leq \kappa \leq \rho_j$ and $1 \leq i \leq l_\kappa$ be chosen as above. Then, the l_j exponential vectors $w_{\kappa i} \exp(\xi_j t) = u_{\kappa i}(D)t^{\kappa-1} \exp(\xi_j t)$ are l_j , P -linearly independent, non-trivial, exponential solutions of S and the degree of $w_{\kappa i} \exp(\xi_j t)$ is $\kappa - 1$. An arbitrary exponential solution $w \exp(\xi_j t)$ of S is P -linearly dependent on the l_j solutions above and on a trivial solution. If the degree of the solution $w \exp(\xi_j t)$ is $\sigma - 1 \leq \rho_j - 1$, then $w \exp(\xi_j t)$ is P -linearly dependent on the $\Sigma_{\kappa=1}^{l_\kappa}$ solutions $w_{\kappa i} \exp(\xi_j t)$ for $1 \leq \kappa \leq \sigma$ and $1 \leq i \leq l_\kappa$, and on a trivial solution.*

Proof. Since $Au_{\kappa i} = (x - \xi_j)^{\kappa} v_{\kappa i}$ and since the P -linear independence of $v_{\kappa 1}, \dots, v_{\kappa l_\kappa} \bmod M : (x - \xi_j)^{\kappa-1}$ implies $v \notin M$, the vectors $w_{\kappa i} \exp(\xi_j t)$ are non-trivial exponential solutions of S according to Theorem 1.41. Let $w \exp(\xi_j t) = u(D)t^{\sigma-1} \exp(\xi_j t)$ be an arbitrary exponential solution of degree $\sigma - 1$ of S . Assuming that $\sigma - 1 \geq \rho_j$, i. e., that $\sigma = \rho_j + \gamma$ where $\gamma \geq 1$, we assert that $w \exp(\xi_j t)$ is the sum of an exponential solution of degree $\leq \rho_j - 1$ and a trivial solution of S . We know that $Au = (x - \xi_j)^\sigma v$ for some $v \in M : (x - \xi_j)^\sigma$ (from Theorem 1.41), and that $M : (x - \xi_j)^\sigma = M : (x - \xi_j)^{\rho_j}$ (from Theorem 1.12). Hence, there exists a vector u' such that $Au' = (x - \xi_j)^{\rho_j} v$ and consequently $u = (x - \xi_j)^\gamma u' + n$ where $n \in N$. This implies that $u(D)t^{\sigma-1} \exp(\xi_j t) = u'(D)(D - \xi_j)^\gamma t^{\sigma-1} \exp(\xi_j t) + n(D)t^{\sigma-1} \exp(\xi_j t) = cu'(D)t^{\rho_j-1} \exp(\xi_j t) + n(D)t^{\sigma-1} \exp(\xi_j t)$ where $c \neq 0$ from which the assertion immediately follows. Now, let $\sigma - 1 \leq \rho_j - 1$ and let V_κ be the $m \times l_\kappa$ matrix whose columns are the vectors $v_{\kappa 1}, \dots, v_{\kappa l_\kappa}$ and let the matrices U_κ and W_κ be similarly defined respectively by the vectors $u_{\kappa 1}, \dots, u_{\kappa l_\kappa}$ and $w_{\kappa 1}, \dots, w_{\kappa l_\kappa}$. To say that $w \exp(\xi_j t) = u(D)t^{\sigma-1} \exp(\xi_j t)$ is P -linearly dependent on the solutions $w_{\kappa i} \exp(\xi_j t)$ for $1 \leq \kappa \leq \sigma$ and $1 \leq i \leq l_\kappa$ means of course that $w \exp(\xi_j t) = (\Sigma_{\kappa=1}^{\sigma} W_\kappa c_\kappa) \exp(\xi_j t)$ where the c_κ 's are constant vectors with l_κ components. A simple computation shows that this is equivalent to saying that $u(D)t^{\sigma-1} \exp(\xi_j t) = \Sigma_{\kappa=1}^{\sigma} (U_\kappa c_\kappa)(D) (t^{\kappa-1} \exp(\xi_j t))$, where $(U_\kappa c_\kappa)(D)$ is the operator vector which arises from the vector $U_\kappa c_\kappa$ by the substitution $x = D$. In order to prove the existence of the σ vectors c_κ we observe that, according to Theorem 1.41, $Au = (x - \xi_j)^\sigma v$ where $v \in M : (x - \xi_j)^\sigma$. Since $v_{\sigma 1}, \dots, v_{\sigma l_\sigma}$ are σ vectors of $M : (x - \xi_j)^\sigma$ which are P -linearly independent $\bmod M : (x - \xi_j)^{\sigma-1}$ and since σ is the P -rank of the factor module $M : (x - \xi_j)^\sigma / M : (x - \xi_j)^{\sigma-1}$, there exists a constant vector c with l_σ components such that $v = V_\sigma c + v'$

where $v' \in M : (x - \xi_j)^{\sigma-1}$. Hence, $(x - \xi_j)^{\sigma} v' = (x - \xi_j) A v''$ which implies that $A(u - (x - \xi_j) v'') = (x - \xi_j)^{\sigma} V_{\sigma} c$. Since also $A U_{\sigma} c = (x - \xi_j)^{\sigma} V_{\sigma} c$, we conclude that $u = (x - \xi_j) v'' + U_{\sigma} c + n$, where $n \in N$. Consequently, $u(D) t^{\sigma-1} \exp(\xi_j t) = v''(D) \exp(\xi_j t) (D t^{\sigma-1}) + (U_{\sigma} c)(D) t^{\sigma-1} \exp(\xi_j t) + n(D) t^{\sigma-1} \exp(\xi_j t)$. The last term is a trivial solution of S and the middle term is a solution which is P -linearly dependent on the solutions $u_{\sigma i}(D) t^{\sigma-1} \exp(\xi_j t)$ for $1 \leq i \leq l_{\sigma}$ of S . Hence, the first term is a solution of S of degree at most $\sigma - 2$ and is not present if $\sigma - 1 = 0$. Hence, if $\sigma - 1 = 0$, the solution $w \exp(\xi_j t)$ is P -linearly dependent on the solutions stated in Theorem 1.42. If $\sigma - 1 > 0$, we can first make the induction hypothesis that the P -linear dependence of $w \exp(\xi_j t)$ on the solutions stated in Theorem 1.42 is proved for degrees $0, 1, \dots, \sigma - 2$. For degree $\sigma - 1$ the same statement then follows immediately from the above remark about the degree of the first term. There remains to be proved that the degree of $w_{\kappa i} \exp(\xi_j t)$ is $\kappa - 1$ and that the l_j solutions of Theorem 1.42 are P -linearly independent. Let $u_{\kappa i}(\xi_j)$ be the constant vector which arises from $u_{\kappa i}$ by the substitution $x = \xi_j$. We assert that the l_{κ} vectors $u_{\kappa 1}(\xi_j), \dots, u_{\kappa l_{\kappa}}(\xi_j)$ are P -linearly independent for each $1 \leq \kappa \leq \rho_j$. To prove this assertion, suppose that $U_{\kappa}(\xi_j) c = 0$ where c is a constant vector with l_{κ} components and where the matrix $U_{\kappa}(\xi_j)$ arises from $U_{\kappa}(x)$ by the substitution $x = \xi_j$. Then $U_{\kappa}(x) c = (x - \xi_j) u'$ and hence $(x - \xi_j) A u' = (x - \xi_j)^{\kappa} V_{\kappa} c$, which implies that $A u' = (x - \xi_j)^{\kappa-1} V_{\kappa} c$. This means that the vector $V_{\kappa} c \in M : (x - \xi_j)^{\kappa-1}$ and hence that $c = 0$ which proves the assertion. The assertion implies that $u_{\kappa i}(\xi_j) \neq 0$ and hence that the degree of $u_{\kappa i}(D) t^{\kappa-1} \exp(\xi_j t)$ is $\kappa - 1$ according to Lemma 1.41. Furthermore,

$$u_{\kappa i}(D) t^{\kappa-1} \exp(\xi_j t) = \exp(\xi_j t) \sum_{g=0}^{\kappa-1} \binom{\kappa-1}{g} u_{\kappa i}^{(g)}(\xi_j) t^{\kappa-1-g}$$

(see the proof of Lemma 1.41) and hence the coefficient of $t^{\kappa-1}$ is the vector $u_{\kappa i}(\xi_j)$. From the above assertion we then conclude that, for each κ , the l_{κ} vectors $u_{\kappa}(D) t^{\kappa-1} \exp(\xi_j t)$ are P -linearly independent and, from the fact that the degree of $u_{\kappa}(D) t^{\kappa-1} \exp(\xi_j t)$ is $\kappa - 1$, we conclude that all the l_j vectors of Theorem 1.42 are P -linearly independent.

With regard to the trivial solutions of S , let n_1, \dots, n_{m-r} be $m - r$ linearly independent generators of N . Clearly, a trivial solution is then P -linearly dependent on the infinite number of P -linearly independent trivial solutions $n_i(D) t^{\kappa} \exp(\xi_j t)$ for $1 \leq i \leq m - r$ and arbitrary κ .

Remark 1.41. If A is non-singular, $r = m$ and S has no trivial solutions. Hence, Theorem 1.42 contains the usual statement about the system S .

namely that, if A is a square non-singular matrix, the number of P -linearly independent exponential solutions $w \exp(\xi_j t)$ of S is equal to the multiplicity l_j of ξ_j as a root of Δ (see [1], Chapter 5).

1. 5. Computation of the solution of S .

Remark 1. 51. The purpose of this section is to show how the solutions of S can be computed by means of the methods of the previous section. The difference between this method and the classical way of computing the solutions (see [1] pp. 168 and 179) is that the matrix A does not have to be restricted to a non-singular square matrix.

The notation will be the same as in 1. 4. Hence A is again an $m \times s$ matrix of rank r whose elements are polynomials of $P[x]$, and the invariants Δ , ϵ , l_j , ρ_j , ξ_j , $\text{Cl}(M)$ etc. all refer to the column space M of A . Again, the letter c denotes elements of P and N is the null space of A .

DEFINITION 1. 51. An $s \times r$ matrix B will be called a ξ_j -multiplier of multiplicity σ of A if $AB = \phi C$, where ϕ is a polynomial of $P[x]$ which has ξ_j as a root of multiplicity σ and where the r columns of C generate $\text{Cl}(M)$.

Since $M \subseteq \text{Cl}(M)$ an $r \times s$ matrix K exists such that $A = CK$ (designated equation m_1) which implies that $ABK = \phi A$ and hence that $BK + F = \phi I$ (designated equation m_2), where F is an $s \times s$ matrix whose columns are elements of N and where I is the $s \times s$ unit matrix. The following theorem properly states how solutions of S can be computed if A is an arbitrary $m \times s$ matrix of rank r .

THEOREM 1. 51. Let B be an ξ_j -multiplier of multiplicity $\sigma \geq \rho_j$ of A . Let the $m \times r$ matrix $U_\kappa(t)$ for $0 \leq \kappa \leq \sigma - 1$ be defined by $U_\kappa \exp(\xi_j t) = B(D)t^\kappa \exp(\xi_j t)$ where $B(D)$ arises from B by the substitution $x = D$. Then, the columns of $U_\kappa \exp(\xi_j t)$ are exponential solutions of S . Conversely, every exponential solution $w \exp(\xi_j t)$ of S is P -linearly dependent on these columns and on a trivial solution.

Proof.

$$A(D)U_\kappa \exp(\xi_j t) = A(D)B(D)t^\kappa \exp(\xi_j t) = C(D)\phi(D)t^\kappa \exp(\xi_j t).$$

Since $\phi(D) = \psi(D)(D - \xi_j)^\sigma$ and $\sigma > \kappa$, $C(D)\phi(D)t^\kappa \exp(\xi_j t) = 0$ and hence every column of $U_\kappa \exp(\xi_j t)$ is an exponential solution of S for $0 \leq \kappa \leq \sigma - 1$. To prove the converse, observe that, according to Theorem 1. 42, every exponential solution of S is the sum of an exponential solution

of degree $\leq \rho_j - 1$ and a trivial solution. Hence we only have to assume that $w \exp(\xi_j t) = u(D)t^{\rho_j-1-\gamma} \exp(\xi_j t)$ is a solution of degree $\rho_j - 1 - \gamma$ of S for $0 \leq \gamma \leq \rho_j - 1$. Then we know that $Au = (x - \xi_j)^{\rho_j-\gamma}v$ where $v \in \text{Cl}(M)$ and hence that $v = Cv'$. Equation m_1 then gives $CKu = (x - \xi_j)^{\rho_j-\gamma}Cv'$ and, since the r columns of C , as generators of $\text{Cl}(M)$, must be linearly independent with respect to $P[x]$, we conclude that $Ku = (x - \xi_j)^{\rho_j-\gamma}v'$. Consequently $BKu = B(x - \xi_j)^{\rho_j-\gamma}v'$ and hence, from equation m_2 , $\phi u = B(x - \xi_j)^{\rho_j-\gamma}v' + Fu$. Since $\phi = (x - \xi_j)^\sigma \psi$ and $\sigma \geq \rho_j$, $Fu = (x - \xi_j)^{\rho_j-\gamma}n$, where $n \in N$, since $Fu \in N$ and a null space is always a closed module. We conclude that, if $\sigma = \rho_j + h$ for $h \geq 0$, $(x - \xi_j)^{h+\gamma}\psi u = Bv' + n$. This gives rise to the relationship $u(D)\psi(D)(D - \xi_j)^{h+\gamma}t^{\rho_j+h-1} \exp(\xi_j t) = B(D)v'(D)t^{\rho_j+h-1} \exp(\xi_j t) + n(D)t^{\rho_j+h-1} \exp(\xi_j t)$. The last term of this relationship is a trivial solution of S . The first term is P -linearly dependent on the columns of $U_\kappa \exp(\xi_j t)$ for $0 \leq \kappa \leq \rho_j + h - 1 = \sigma - 1$ as is seen by first expanding $v'(D) = \sum_i c_i D^i$ and then applying this operator to $t^{\rho_j+h-1} \exp(\xi_j t)$. For the left-hand side we write $u(D)\psi(D)(D - \xi_j)^{h+\gamma}t^{\rho_j+h-1} \exp(\xi_j t) = cu(D)\psi(D)t^{\rho_j-\gamma-1} \exp(\xi_j t)$, where $c \neq 0$. Since $\psi(D) = \sum d_i (D - \xi_j)^i$, where $d_i \in P$ and $d_0 \neq 0$ because $\psi(\xi_j) \neq 0$, application of the operator $\psi(D)$ to $t^{\rho_j-\gamma-1} \exp(\xi_j t)$ shows that $cu(D)\psi(D)t^{\rho_j-\gamma-1} \exp(\xi_j t) = dw \exp(\xi_j t) + w' \exp(\xi_j t)$ where d is the non-zero element $d_0 c$ of P and where $w \exp(\xi_j t)$ is the given exponential solution while $w' \exp(\xi_j t)$ is an exponential solution of degree $< \rho_j - \gamma - 1$ of S . Hence, if the degree of $w \exp(\xi_j t)$ is zero, i. e., if $\rho_j - 1 - \gamma = 0$, the term $w' \exp(\xi_j t)$ does not occur and the theorem is proved. For arbitrary degree $\rho_j - 1 - \gamma$, we first make the induction hypothesis that the theorem is proved for degrees $0, 1, \dots, \rho_j - 2 - \gamma$, and the theorem then follows for degree $\rho_j - 1 - \gamma$ from the remark about the degree of $w' \exp(\xi_j t)$.

There always exist matrices B which are ξ_j -multipliers of A of multiplicity ρ_j simultaneously for all roots ξ_j of ϵ , for, if C is any matrix whose columns generate $\text{Cl}(M)$, then ϵC is a matrix whose columns are vectors of M . Hence, there exists a matrix B such that $AB = \epsilon C$ and this B clearly satisfies the requirement.

Remark 1.52. In the classical method of computing the solutions of S it is assumed that A is a square non-singular matrix and hence its adjoint H is then well defined. The substitution $x = \xi_j$ is made in the derivatives of order $0, 1, \dots, l_j - 1$ of $\exp(xt)H(x)$ with respect to x (see [1], pp. 168 and 179). However, H is then a ξ_j -multiplier of A of multiplicity $l_j \geq \rho_j$ simultaneously for all roots ξ_j of ϵ , because $AH = \Delta I$ and the columns of

the unit matrix I generate the whole vector space V , which is $\text{Cl}(M)$, since M is dense in this case. Furthermore

$$\begin{aligned} B(D)t^k \exp(\xi_j t) &= \exp(\xi_j t) \sum_{i=0}^k \binom{k}{i} B^{(i)}(\xi_j) t^{k-i} \\ &= \exp(\xi_j t) [(D_x + t)^k B(x)]_x = \xi_j = [D_x^k \exp(xt) \cdot B(x)]_x = \xi_j \end{aligned}$$

which shows that the classical statement, mentioned above is contained in Theorem 1.51.

1.6. Trivial solutions in terms of the row space. According to Theorem 1.42 the properties of the solution of S are determined by two column space invariants, namely Δ and ϵ , and by the trivial solutions. The trivial solutions are defined in terms of the null space N which is not an invariant of the column space M . The rows of A , however, generate the row space M' which is a module of the m -dimensional vector space V' consisting of the row vectors whose m components are elements of $P[x]$. Since the Δ and ϵ of M and M' can be computed in the same way from the sub-determinants of A , the Δ and ϵ of M and M' are the same. Hence, we can as well say that the properties of S are determined by two row space invariants and by the trivial solutions. We shall now show how the trivial solutions could have been defined invariantly in terms of the row space.

Remark 1.61. The purpose of this section is to give the basis for the notion of trivial solution for a system of partial differential equations which will be discussed in MSV. Since, if the scalar domain is $\mathfrak{s} = P[x_1, \dots, x_n]$, the invariants Δ and ϵ can not be computed any more as sub-determinants of A , a column space and a row space of the same matrix may then very well have different invariants. It will be shown in MSV that if $\mathfrak{s} = P[x_1, \dots, x_n]$, the invariants of the row space M' , and not of the column space M , are essential for the system of partial differential equations determined by the matrix. In particular, trivial solutions then have to be defined in terms of M' . It will become clear from MSV how the theory of 1.4 could have been developed, less simply but yet completely, in terms of the row space M' . This is no wonder since, if $\mathfrak{s} = P[x]$, the fact that the invariants of M and M' can be computed by means of sub-determinants of A , implies that M and M' then have the same invariants.

The system S is given by m differential operators which arise from the m rows of A by the substitution $x = D$. Since every vector of M' is a linear combination of the rows of A , the system of differential equations $M'(D)$, which is given by the infinite number of differential operators which arise

from all the rows of M' by the substitution $x = D$, has the same solutions as S . If $\text{Cl}(M')(D)$ arises from $\text{Cl}(M')$ in the same way as $M'(D)$ from M' , then, since $M' \subseteq \text{Cl}(M')$, the solutions of $\text{Cl}(M')(D)$ are contained among those of $M'(D)$, i. e., among those of S . The following theorem shows how the trivial solutions of S could have been defined intrinsically in terms of M' .

THEOREM 1.61. *An exponential solution $w \exp(\xi t)$ of S is a trivial solution of S in the sense of Definition 1.41, if and only if it is a solution of $\text{Cl}(M')(D)$.*

Proof. Let $w \exp(\xi t)$ be a trivial exponential solution of S , i. e., let $w \exp(\xi t) = n(D)t^k \exp(\xi t)$ where $An = 0$. Let $v'(D) \in \text{Cl}(M')(D)$, i. e., v' is a row vector of $\text{Cl}(M')$. Since ϵ is the elementary divisor of M' , we conclude that $\epsilon v' \in M'$ and hence there exists a row vector $s' \in V$ such that $s'A = \epsilon v'$. This implies that $\epsilon v'n = 0$ and hence that $v'n = 0$ from which it follows that $v'(D)w \exp(\xi t) = v'(D)n(D)t^k \exp(\xi t) = 0$. Consequently, every trivial solution is a solution of $\text{Cl}(M')(D)$. Conversely, if $w \exp(\xi t) = u(D)t^k \exp(\xi t)$ is a solution of $\text{Cl}(M')(D)$, it is also a solution of $M'(D)$, and hence all that remains to be proved is that this solution is trivial. Let F be any matrix whose rows generate $\text{Cl}(M')$; then, since $F(D)u(D)t^k \exp(\xi t) = 0$, we have $Fu = (x - \xi)^k v$. Since the rows of F generate a closed module, i. e., a module whose Δ is a non-zero element of P , the determinantal factor of highest rank of F , i. e., Δ , is an element of P . Hence, the columns of F also generate a closed module, which implies the existence of a vector $s \in V$ such that $Fs = v$. Consequently, according to Theorem 1.41, $u(D)t^k \exp(\xi t)$ is a trivial solution of $F(D)y = 0$, i. e., $u(D)t^k \exp(\xi t) = n(D)t^k \exp(\xi t)$ where $Fn = 0$. Since the rows of A are vectors of M' and hence of $\text{Cl}(M')$, we can take for F a matrix whose first m rows are those of A . We then conclude that $An = 0$ and hence that $w \exp(\xi t)$ is a trivial solution of S .

Part II. Structure of Modules.

2.1. Operations and the notion of primary. Let \mathfrak{s} be a commutative ring with unit element and zero element ω . The elements of \mathfrak{s} , called scalars, will be denoted by lower case Greek letters and the ideals of \mathfrak{s} by lower case German letters. Let V be an additive abelian group with zero element 0 for which \mathfrak{s} is the left operator domain and for which the unit element of \mathfrak{s} is the unit operator. The elements of V , called vectors, will be denoted by lower case Roman letters and the \mathfrak{s} -modules of V , called modules, by capital Roman letters. Consequently, a module is a subset of V which is closed under

scalar multiplication and vector-subtraction. We assume furthermore that for the modules of V the ascending chain condition holds, i. e., that each module has a finite number of generators. For the terms used so far, see [9] and [11]. In accordance with [12], p. 690, we call V a *Noetherian vector space* with \mathfrak{s} as scalar domain, or more briefly, a *Noetherian vector space over \mathfrak{s}* . If, without further explanation, a lower case Greek letter, a lower case German letter, a lower case Roman letter or a capital Roman letter is used, these letters will always respectively denote a scalar, an ideal, a vector and a module.

Remark 2.11. The assumptions made in the previous paragraph are indicated by $[0, I]$ in [5]. The vector space of Part I is an example of a Noetherian vector space over $P[x]$ as scalar domain.

For the consistency of the following definitions, see [5] Section 2 or [6] Section 2.

The intersection $M \cap L$ is the module which is the set theoretic intersection of M and L . It is convenient for 2.3 to say, as in [5], that a void collection of modules has V as intersection. The sum (M, L) is the module which is generated by the set theoretic union $M \cup L$ and hence consists of the vectors $v = m + l$ where $m \in M$ and $l \in L$. In the same way, (v_1, \dots, v_s) denotes the module generated by v_1, \dots, v_s and $(\alpha_1, \dots, \alpha_s)$ the ideal generated by $\alpha_1, \dots, \alpha_s$. The quotient $M : L$ is the ideal of \mathfrak{s} which consists of the scalars α such that $\alpha l \in M$ for all $l \in L$. Here, L can be an arbitrary set of vectors, not necessarily a module. The quotient $M : \mathfrak{b}$ is the module which consists of all vectors v such that $\beta v \in M$ for all $\beta \in \mathfrak{b}$. Again, \mathfrak{b} can be any set of scalars, not necessarily an ideal. The product $\mathfrak{b}M$ is the module which is generated by all products βm where $\beta \in \mathfrak{b}$ and $m \in M$. Again, \mathfrak{b} may be an arbitrary set of scalars. The rules which combine intersections, sums, quotients and products are the same as for ideals (see [5] Section 2 or [6] Section 2).

DEFINITION 2.11. The fundamental ideal \mathfrak{f} of M is the quotient $M : V$. The radical \mathfrak{r} of M is the radical of \mathfrak{f} .

The radical of an ideal is as usual the ideal consisting of the scalars of which a power lies in the ideal.

Remark 2.12. The term "fundamental ideal" in definition 2.11 means the same as the terms "shadow" in [5] and "essential ideal" in [6]. The term "shadow" is dropped since it does not indicate that the structure of \mathfrak{f} is of "fundamental" importance for M . The reason for changing the terminology of [6] is to keep the German letter e free to denote "elementary divisor."

From now on, the annihilating ideal of M will always be denoted by α , i. e., $\alpha(M) = 0 : M$. Since $M_1 \subseteq M_2$ implies $f(M_1) \subseteq f(M_2)$ and $r(M_1) \subseteq r(M_2)$, it is always true that $\alpha(V) \subseteq f(M)$ and $r(0) \subseteq r(M)$. The following definition and theorem occur in [5], Section 4, and in [6], Section 4.

DEFINITION 2.12. A module Q is called *primary* if $\alpha v \in Q$ and $\alpha \notin r(Q)$ implies $v \in Q$.

THEOREM 2.11. The fundamental ideal of a primary module is a primary ideal and hence its radical is a prime ideal. A module Q is \mathfrak{p} -primary, i. e., is primary and has \mathfrak{p} as radical, if and only if the following three conditions are satisfied: 1) $\alpha v \in Q$ and $v \notin Q$ implies $\alpha \in \mathfrak{p}$; 2) $f(Q) \subseteq \mathfrak{p}$; 3) $\alpha \in \mathfrak{p}$ implies that, for some integer κ , $\alpha^\kappa \in f(Q)$.

Since, if Q is \mathfrak{p} -primary, $\mathfrak{p} \supseteq f(Q) \supseteq \alpha(V)$ and since the ascending chain condition holds in $\mathfrak{s}/\alpha(V)$ (see [5], Theorem 4), there is a smallest integer ρ such that $\mathfrak{p}^\rho \subseteq f(Q)$. This ρ is called the *exponent* of Q . If Q is \mathfrak{p} -primary and $f(Q)$ is prime, i. e., if $f(Q) = \mathfrak{p}$, and $\rho = 1$, then Q is called a *prime module* or \mathfrak{p} -*prime*. The following lemma is proved in [5], Section 4.

LEMMA 2.11. If Q is \mathfrak{p} -primary, $\mathfrak{b} \subseteq f(Q)$ and $M \subseteq Q$, then $Q : \mathfrak{b}$ is a \mathfrak{p} -primary module and $Q : M$ a \mathfrak{p} -primary ideal.

If $M \subseteq L$ and if M is \mathfrak{p} -primary when M is considered as a sub-module of L instead of V , then M is said to be \mathfrak{p} -primary in L . Since the fundamental ideal of M as a sub-module of L is $M : L$ and not $M : V$, it may be that M is \mathfrak{p} -primary in L but not in V . However, if M is \mathfrak{p} -primary in V , it is certainly \mathfrak{p} -primary in L (if $M \neq L$) as is implied by Lemma 2.12, which follows from the three conditions of Theorem 2.11.

LEMMA 2.12. Let Q be \mathfrak{p} -primary and let L be any module, not contained in Q . Then $Q \cap L$ is \mathfrak{p} -primary in L and the \mathfrak{p} -primary fundamental ideal of $Q \cap L$ as a sub-module of L is $Q : L$.

Remark 2.13. Lemma 2.12, which does not occur in [5] and [6], will be used frequently in that part of the theory of polynomial matrices in which the imbedding vector space V has to be replaced by an isolated component of the row space or column space of the matrix. See also Lemma 2.51 of this paper.

2.2. Noether decompositions.

DEFINITION 2.21. An intersection $M = Q_1 \cap \cdots \cap Q_n$ is called a *Noether decomposition* (N. D. for short) of M if the following three conditions are

satisfied: 1) Q_j is \mathfrak{p}_j -primary for $j = 1, \dots, h$; 2) the prime ideals $\mathfrak{p}_1, \dots, \mathfrak{p}_h$ are all different; 3) the intersection is irredundant.

Remark 2.21. The term "Noether decomposition" is used in preference to the term "Lasker reduction" of [5] to conform to the expression Noetherian vector space.

The following theorem is proved in [5], Section 6, and in [6], Sections 4 and 5.

THEOREM 2.21. *Every module M has an N. D. The radicals of the primary components are the same for all N. D.'s of M and are called the associated primes (a. p. for short) of M . The intersection of the a. p.'s of M is the radical of M .*

A primary component of M is a primary module which occurs in some N. D. of M and a complete set of primary components of M is a set of primary components which together form an N. D. of M . The intersection of a subset of a complete set of primary components of M is called a component of M . From Lemma 2.12 the following lemma can be easily derived.

LEMMA 2.21. *Let $M = Q_1 \cap \dots \cap Q_h$ be an N. D. of M , where Q_j is \mathfrak{p}_j -primary for $1 \leq j \leq h$. Let C be the component $C = Q_1 \cap \dots \cap Q_\kappa$, where $1 \leq \kappa \leq h$, and let $Q_{\kappa+i}^* = C \cap Q_{\kappa+i}$, where $1 \leq i \leq h - \kappa$. Then $M = Q_{\kappa+1}^* \cap \dots \cap Q_h^*$ is an N. D. of M as a submodule of C and the a. p.'s of M as a submodule of C are $\mathfrak{p}_{\kappa+1}, \dots, \mathfrak{p}_h$.*

Remark 2.22. Lemma 2.21 does not occur in [5] and [6] and is important for the same reason as Lemma 2.12.

A component of M is called an *isolated component* if every a. p. of M which is contained in an a. p. of that component is itself an a. p. of that component. An isolated component of M is uniquely determined by its a. p.'s as is proved in [5], Section 6, and in [6], Section 5. A module is called univalent ("monotypic" in [5]) if its radical is a maximal ideal of \mathfrak{S} . Since a univalent module can clearly have only one a. p., namely its own radical, a univalent module is always primary. Finally, an arbitrary module and an ideal c are said to be *relatively prime* if $M : c = M$. It is shown on page 254 of [5] that M and c are relatively prime if and only if no a. p. of c is contained in an a. p. of M .

Remark 2.23. According to the above, if \mathfrak{p} is a prime ideal, $M : \mathfrak{p} \supset M$ (the symbol \supset again denotes *proper* inclusion) if and only if \mathfrak{p} is contained in one of the a. p.'s of M . Since in the scalar domain $P[x]$ of Part I all

prime ideals are maximal, in that particular case $M : \mathfrak{p} \supset M$ if and only if \mathfrak{p} is an a. p. of M . Hence, the non-zero a. p.'s of Part I are those a. p.'s of an N. D. of M which are different from the zero ideal.

2.3. Closures of a module. Throughout this section, let \mathfrak{C} denote a multiplicatively closed subset of \mathfrak{A} which does not contain ω .

DEFINITION 2.31. *The \mathfrak{C} -closure $M(\mathfrak{C})$ of a module M is the module which consists of the vectors v for which there exists a scalar $\gamma \in \mathfrak{C}$ such that $\gamma v \in M$. If $M(\mathfrak{C}) = M$, we say that M is \mathfrak{C} -closed, and if $M(\mathfrak{C}) = V$ that M is \mathfrak{C} -dense.*

Clearly, $M \subseteq M(\mathfrak{C})$ and, if $M_1 \subseteq M_2$, then $M_1(\mathfrak{C}) \subseteq M_2(\mathfrak{C})$.

Remark 2.31. If, in the scalar domain $P[x]$ of Part I, we take for \mathfrak{C} the set of all the non-zero scalars, the \mathfrak{C} -closure becomes the closure $\text{Cl}(M)$ of Definition 1.12. This particular closure occurs in [6] Section 3 for general m -dimensional vector spaces. As will be shown in MSV, the n different closures of a polynomial matrix in n variables, as defined in [13] Section 3, are closures in the sense of Definition 2.31.

The important properties of closures are enumerated in the following lemma whose proof can be found in [5] Section 7.

LEMMA 2.31. *The \mathfrak{C} -closure $M(\mathfrak{C})$ is the smallest \mathfrak{C} -closed module which contains M and is the isolated component of M whose a. p.'s contain no element of \mathfrak{C} . Furthermore, $[M_1 \cap M_2](\mathfrak{C}) = M_1(\mathfrak{C}) \cap M_2(\mathfrak{C})$, from which it follows that the intersection of a finite number of \mathfrak{C} -closed modules is \mathfrak{C} -closed, that the intersection of a finite number of \mathfrak{C} -dense modules is \mathfrak{C} -dense, and the irredundant intersection of a \mathfrak{C} -closed module and a \mathfrak{C} -dense module is neither \mathfrak{C} -closed nor \mathfrak{C} -dense.*

By considering all possible \mathfrak{C} -closures of M we obtain all possible isolated components of M (see [5] Theorem 16) together with V . Namely, if every a. p. of M has an element in common with \mathfrak{C} , then M is \mathfrak{C} -dense, according to the convention that the intersection of a void collection of modules is V . If \mathfrak{C} is the complement of a prime ideal \mathfrak{p} of \mathfrak{A} , we write $M(\mathfrak{p})$, \mathfrak{p} -closure and \mathfrak{p} -closed instead of $M(\mathfrak{C})$, \mathfrak{C} -closure and \mathfrak{C} -closed, following the custom of [5] for modules and [14] for ideals. Hence, $M(\mathfrak{p})$ is the isolated component of M whose a. p.'s are contained in \mathfrak{p} . It follows that M is \mathfrak{p} -closed if and only if all a. p.'s of M are contained in \mathfrak{p} and is \mathfrak{p} -dense if and only if none of its a. p.'s are contained in \mathfrak{p} . In particular, a \mathfrak{p} -primary module is \mathfrak{p} -closed.

Remark 2.32. If V is the vector space of Part I, $\text{Cl}(M) = M((\omega))$,

where (ω) is the zero ideal of $P[x]$; and closed and dense then mean (ω) -closed and (ω) -dense. Hence, in that case, a module M is dense if and only if (ω) is not an a. p. of M . If, however, M is not dense, (ω) is an a. p. of M and $\text{Cl}(M)$ is then the isolated primary component of M whose a. p. is (ω) . The occurrence of the zero ideal as an a. p. does not happen in ideal theory since ideals of integral domains, considered as modules, are always (ω) -dense.

We now discuss the notion of quotient module which is related to that of \mathfrak{C} -closed module and is needed for the theory of length of a primary module. The *quotient module* $V_{\mathfrak{C}}$ consists of classes of ordered pairs $\{v, \gamma\}$ where $v \in V$, $\gamma \in \mathfrak{C}$ and two pairs $\{v_1, \gamma_1\}$ and $\{v_2, \gamma_2\}$ belong to the same class if $\gamma_2 v_1 = \gamma_1 v_2$. According to [5] Section 13, addition and scalar multiplication can be defined in the usual way in $V_{\mathfrak{C}}$, which makes $V_{\mathfrak{C}}$ into a Noetherian vector space over $\mathfrak{s}_{\mathfrak{C}}$ as scalar domain. Here, $\mathfrak{s}_{\mathfrak{C}}$ arises from \mathfrak{s} as $V_{\mathfrak{C}}$ arises from V . By forming quotients, the module $M_{\mathfrak{C}} = T(M)$ of $V_{\mathfrak{C}}$ corresponds to the module M of V . In the same way, the ideal $\mathfrak{b}_{\mathfrak{C}} = A(\mathfrak{b})$ of $\mathfrak{s}_{\mathfrak{C}}$ corresponds to the ideal \mathfrak{b} of \mathfrak{s} . The following lemma is proved in [5] Section 13.

LEMMA 2.32. *The correspondence T is a one-to-one correspondence between the \mathfrak{C} -closed modules of V and the modules of $V_{\mathfrak{C}}$ and A is a one-to-one correspondence between the \mathfrak{C} -closed ideals of \mathfrak{s} and the ideals of $\mathfrak{s}_{\mathfrak{C}}$. Both these correspondences are isomorphisms with respect to the quotient operation, the intersection operation \cap , and the inclusion relation \subset . It follows that, if M is \mathfrak{C} -closed, $\mathfrak{f}(T(M)) = A(\mathfrak{f}(M))$ and $\mathfrak{r}(T(M)) = A(\mathfrak{r}(M))$. Furthermore M is \mathfrak{p} -primary if and only if $T(M)$ is $A(\mathfrak{p})$ -primary, in which case M and $T(M)$ have the same exponents.*

Again, if \mathfrak{C} is the complement of a prime ideal \mathfrak{p} of \mathfrak{s} , we write $V_{\mathfrak{p}}$, $\mathfrak{s}_{\mathfrak{p}}$ etc. instead of $V_{\mathfrak{C}}$, $\mathfrak{s}_{\mathfrak{C}}$.

2.4. Length of primary modules. The following definition occurs for ideals in [10] Section 23.

DEFINITION 2.41. *A sequence $Q_0 \subset Q_1 \subset \cdots \subset Q_{l-1} \subset V$ is called a \mathfrak{p} -primary composition sequence from Q_0 to V of length l if 1) Q_i is \mathfrak{p} -primary for $0 \leq i \leq l-1$ and 2) it is impossible to insert a \mathfrak{p} -primary module Q between any two consecutive members $Q_i \subset Q_{i+1}$ such that $Q_i \subset Q \subset Q_{i+1}$.*

Let Q be a \mathfrak{p} -primary module. Since Q is \mathfrak{p} -closed, $T(Q)$ is a primary $A(\mathfrak{p})$ module of $V_{\mathfrak{p}}$. Since $A(\mathfrak{p})$ is a maximal prime ideal of $\mathfrak{s}_{\mathfrak{p}}$ (see [5] Section 14), $T(Q)$ is even univalent and hence the factor module $V_{\mathfrak{p}}/T(Q)$ is a Noetherian vector space over $\mathfrak{s}_{\mathfrak{p}}$ as scalar domain whose zero module is univalent. This implies (see [5] Section 11) that $V_{\mathfrak{p}}/T(Q)$ has an ordinary

composition sequence whose factors (i. e., the factor modules formed by consecutive elements) are all isomorphic with $\mathfrak{S}_p/A(p)$. These isomorphisms are operator isomorphisms with respect to \mathfrak{S}_p as operator domain. The theory of composition sequences, applied to $V_p/T(Q)$, then gives, by the use of Lemma 2.32, the following theorem.

THEOREM 2.41. *For every p -primary module Q we can draw a p -primary composition sequence $Q \subset Q_1 \subset \cdots \subset Q_{l-1} \subset V$ from Q to V . All these sequences have the same length l , called the length of Q . Any sequence $Q \subset H_1 \subset \cdots \subset H_s \subset V$ where the H_i 's are p -primary can be refined to a p -primary composition sequence. If Q is univalent, a p -primary composition sequence from Q to V is an ordinary composition sequence, i. e., no module, p -primary or not, can be inserted between two consecutive elements. In that case the factors are all operator isomorphic with \mathfrak{S}/p with respect to \mathfrak{S} as operator domain.*

Remark 2.41. In Theorem 1.13 it was shown that, for the vector space of Part I, a composition sequence can be drawn from any module M , primary or not, to any other module N where $M \subseteq N \subseteq \text{Cl}(M)$. This is caused by the fact that all non-zero prime ideals of $P[x]$ are maximal and hence all primary components of M , except $\text{Cl}(M)$, are univalent (see [5] Theorem 24).

2.5. Elementary divisors and lengths of arbitrary modules. Let p be an a. p. of M and let $M(p)$ again be the p -closure of M , i. e., the isolated component of M whose a. p.'s are contained in p . Let $M'(p)$ be the isolated component of M whose a. p.'s are properly contained in p . Then, if Q is any p -primary component of M , $M(p) = M'(p) \wedge Q$. The following lemma then follows immediately from Lemmas 2.12 and 2.21.

LEMMA 2.51. $M(p) \subset M'(p)$ and $M(p)$ is p -primary in $M'(p)$, namely it is the isolated p -primary component of M as a sub-module of $M'(p)$. The p -primary fundamental ideal of $M(p)$ as a sub-module of $M'(p)$ is $M(p) : M'(p) = Q : M'(p)$.

DEFINITION 2.51. The p -primary ideal $e(p) = M(p) : M'(p)$ is called the p -elementary divisor of M . The exponent $\rho(p)$ of $e(p)$ is called the p -exponent of M . The length $l(p)$ of $M(p)$ as a primary sub-module of $M'(p)$ is called the p -length of M .

Remark 2.51. The p -elementary divisors for $p \neq (\omega)$ are the analogues of the primary factors of the elementary divisor ϵ of Part I. Let M be a

module of the vector space of Part I and let $M = \text{Cl}(M) \circ Q_1 \circ \cdots \circ Q_h$ be an *N.D.* of M where Q_j is \mathfrak{p}_j -primary and $\mathfrak{p}_j \neq (\omega)$. Then, $M(\mathfrak{p}_j) = \text{Cl}(M) \circ Q_j$, $M'(\mathfrak{p}_j) = \text{Cl}(M)$, and hence $\mathfrak{c}(\mathfrak{p}_j) = Q_j : \text{Cl}(M)$. Since all \mathfrak{p}_j 's are relatively prime for $j = 1, \cdots, h$, we conclude that

$$\begin{aligned} \Pi_{j=1}^h \mathfrak{c}(\mathfrak{p}_j) &= \bigcap_{j=1}^h \mathfrak{c}(\mathfrak{p}_j) = \bigcap_{j=1}^h (Q_j : \text{Cl}(M)) = (\bigcap_{j=1}^h Q_j) : \text{Cl}(M) \\ &= (\text{Cl}(M) \circ \bigcap_{j=1}^h Q_j) : \text{Cl}(M) = M : \text{Cl}(M) = \epsilon, \end{aligned}$$

which proves the assertion. The interpretation of $\rho(\mathfrak{p}_j)$, given in Theorem 1.12, is in the general case represented by the fact that $\rho(\mathfrak{p}_j)$ is the exponent of $M(\mathfrak{p}_j) : M'(\mathfrak{p}_j)$, i. e., that $\mathfrak{p}_j^{\rho(\mathfrak{p}_j)+h} M'(\mathfrak{p}_j) \subseteq M(\mathfrak{p}_j)$ for $h \geq 0$ and not for $h < 0$. Furthermore, $l(\mathfrak{p}_j)$ is the length of $\text{Cl}(M) \circ Q_j$ as a primary sub-module of $\text{Cl}(M)$ from which it follows easily that $l(\mathfrak{p}_j)$ is equal to the multiplicity l_j of the norm $\Delta = \Pi_{j=1}^h \pi^{l_j}$ of M (see Definition 1.13). Finally, $M'(\mathfrak{p}_j) = \text{Cl}(M)$ makes it clear that, for a polynomial matrix in n variables, the *different* closures $M'(\mathfrak{p})$ for the a. p.'s of the column space (or for other applications of the row space) are the analogues of the one closure $\text{Cl}(M)$ of a polynomial matrix in one variable.

2.6. Systems of linear equations. The following criterion is the abstract formulation of the known criteria for the solvability of linear equations over different types of scalar domains \mathfrak{s} (see the **Introduction**).

Criterion of Lengths. Let M_1 and M_2 be two modules where $M_1 \subseteq M_2$. Then, $M_1 = M_2$ if and only if M_1 and M_2 have the same a. p.'s and for each a. p. \mathfrak{p} , the same \mathfrak{p} -length.

Proof. It is obvious that if $M_1 = M_2$, the a. p.'s and \mathfrak{p} -lengths of M_1 and M_2 are the same. Hence, we assume that $M_1 \subseteq M_2$ and that M_1 and M_2 have the same a. p.'s and \mathfrak{p} -lengths. Let \mathfrak{p} be an a. p. of M_1 and M_2 such that $M_1(\mathfrak{p})$ and hence $M_2(\mathfrak{p})$ have only one a. p., i. e., let \mathfrak{p} be an isolated a. p. of M_1 and M_2 . We then conclude from 2.3 that $M_1(\mathfrak{p}) \subseteq M_2(\mathfrak{p})$ and from Lemma 2.51 that $M_1(\mathfrak{p})$ and $M_2(\mathfrak{p})$ are both \mathfrak{p} -primary in $M'(\mathfrak{p})$ and from Definition 2.51 that they have the same length as \mathfrak{p} -primary sub-modules of $M'(\mathfrak{p})$. Hence $M_1(\mathfrak{p}) = M_2(\mathfrak{p})$ since, otherwise, it would follow immediately from Theorem 2.41 that the length of $M_1(\mathfrak{p})$ is greater than the length of $M_2(\mathfrak{p})$. We then make the induction hypothesis that $M_1(\mathfrak{p}) = M_2(\mathfrak{p})$ has been proved for all a. p.'s for which $M_1(\mathfrak{p})$ and hence $M_2(\mathfrak{p})$ have at most $\kappa - 1$ a. p.'s. Let \mathfrak{p} then be such that $M_1(\mathfrak{p})$ and $M_2(\mathfrak{p})$ have κ a. p.'s. Since $M'_1(\mathfrak{p})$ and $M'_2(\mathfrak{p})$ have at most $\kappa - 1$ a. p.'s, it follows easily from the

induction hypothesis that they are equal, say $M' = M'_1(p) = M'_2(p)$. Then, as before, $M_1(p)$ and $M_2(p)$ are p -primary in M' and have the same length as p -primary sub-modules of M' ; and $M_1(p) \subseteq M_2(p)$. Hence $M_1(p) = M_2(p)$ for all a. p.'s of M , which proves $M_1 = M_2$.

The connection between the criterion of lengths and systems of linear equations is established by considering the m -dimensional Noetherian vector space over \mathfrak{s} which consists of the column vectors whose m components are elements of \mathfrak{s} . The columns of an $m \times s$ matrix $A = (\alpha_{ij})$, where $\alpha_{ij} \in \mathfrak{s}$ and $i = 1, \dots, m$ and $j = 1, \dots, s$, generate a module M , called the column space of A ; and the $e(p)$, $\rho(p)$ and $l(p)$ of A are defined as those of the column space.

Criterion of Solvability. Let $\sum_{j=1}^s \alpha_{ij} z_j = \gamma_i$, where $i = 1, \dots, m$ and α_{ij} and γ_i are elements of \mathfrak{s} , be a system of m linear equations for the s unknowns z_j . Then, this system has a solution $z_j \in \mathfrak{s}$ ($j = 1, \dots, s$) if and only if the $m \times s$ matrix $A = (\alpha_{ij})$ and the augmented $m \times (s+1)$ matrix $B = (\alpha_{ij}, \gamma_i)$ have the same a. p.'s and for each a. p. p , the same p -length.

Proof. Let M be the column space of A and g the column vector whose components are $\gamma_1, \dots, \gamma_m$. The system has a solution if and only if $\gamma \in M$, i. e., if and only if $M = L$, where L is the module (M, γ) . Since clearly $M \subseteq L$ and L is the column space of B , the criterion of solvability is an immediate corollary of the criterion of lengths.

2.7. Rank and $\text{Cl}(M)$ of a module. In this section, except in Lemma 2.71, V will denote a Noetherian vector space over \mathfrak{s} as scalar domain whose zero-module (0) is primary. We want to show that for such a module a notion of rank and $\text{Cl}(M)$ exists which has the same properties as in 1.2 and which is contained in the more general notion of length of a primary composition sequence.

The radical of the zero-module will be denoted by r ; hence r is a prime ideal. The radical of an arbitrary module $M \neq (0)$ will be denoted by $r(M)$.

DEFINITION 2.71. The closure $\text{Cl}(M)$ of a module M is the r -closure of M . To say that M is closed or dense means that M is respectively r -closed or r -dense.

It follows from the general theory of 2.3 that $\text{Cl}(M)$ has the properties which are described in 1.2 and in [6], Section 3. Since $r \subseteq r(M)$, either M is dense, i. e., $\text{Cl}(M) = V$, or $\text{Cl}(M)$ is an isolated primary component of

M whose a. p. is r .^{*} In particular, a primary module is either closed or dense (compare with [6], Theorem 4.2).

Since (0) is r -primary as a sub-module of V , the zero-module (0) is also r -primary in any module M (see Lemma 2.12).

DEFINITION 2.72. *The rank r of a module M is the length of (0) as an r -primary sub-module of M .*

The fact that rank and $\text{Cl}(M)$ behave as in 1.2 follows from the following lemma.

LEMMA 2.71. *Let V be an arbitrary Noetherian module with arbitrary zero-module and let Q be a p -primary module of V . Then, if $Q \subseteq M$, the p -closure $M(p)$ is the largest module of V which satisfies the two conditions: 1) $M \subseteq M(p)$; 2) the length of Q as a p -primary sub-module of M is equal to the length of Q as a p -primary sub-module of $M(p)$.*

Proof. The lemma follows immediately from the fact that, in the notation of 2.3, $M(p)$ is the largest module containing M such that $T(M) = T(M(p))$.

If we apply Lemma 2.71 to the case that Q is the r -primary zero-module of our Noetherian vector space V , all statements of 1.2 about rank and $\text{Cl}(M)$ are again seen to hold. In particular, if $M_1 \subseteq M_2$ and if M_1 and M_2 have the same rank, then $\text{Cl}(M_1) = \text{Cl}(M_2)$. We can now reformulate the criterion of lengths, using r instead of the r -length of M .

THEOREM 2.71. *Let V be a Noetherian vector space whose zero-module is primary and let $M_1 \subseteq M_2$ be two modules of V . Then, $M_1 = M_2$, if and only if M_1 and M_2 have the same rank, the same a. p.'s different from r and for each such a. p. p , the same p -length.*

Proof. The proof follows immediately from the fact that, if $M_1 \subseteq M_2$ and M_1 and M_2 have the same rank, then $\text{Cl}(M_1) = \text{Cl}(M_2)$. Hence, either M_1 and M_2 do not have r as an a. p. and Theorem 2.71 immediately reduces to the criterion of lengths; or r is an a. p. of M_1 and M_2 , in which case M_1 and M_2 have the same isolated r -primary component $\text{Cl}(M_1) = \text{Cl}(M_2)$ and hence also the same r -length.

In order to reformulate also the criterion of solvability, let V be the m -dimensional Noetherian vector space consisting of the column vectors whose m components are elements of \mathfrak{K} . We again assume that the zero-module of V is r -primary and define the rank of a matrix A as the rank of its column

space. The following theorem is then an immediate corollary of the criterion of solvability.

THEOREM 2.72. *Let $\sum_{j=1}^s \alpha_{ij} z_j = \gamma_i$, where $i = 1, \dots, m$ and α_{ij} and γ_i are elements of \mathfrak{S} , be a system of m linear equations for the s unknowns z_j . Then, this system has a solution $z_j \in \mathfrak{S}$ ($j = 1, \dots, s$) if and only if the $m \times s$ matrix $A = (\alpha_{ij})$ and the augmented $m \times (s + 1)$ matrix $B = (\alpha_{ij}, \gamma_i)$ have the same rank, the same a. p.'s different from \mathfrak{r} and for each such a. p. the same \mathfrak{p} -length.*

Remark 2.71. In the vector space of Part I, the zero-module is an (ω) -prime module and hence Theorem 2.72 applies. Since the a. p.'s of a module M which are different from (ω) are then the irreducible factors of the norm $\Delta = \prod_{j=1}^h \pi_j^{l_j}$ of M and since the multiplicity l_j is the (π_j) -length of M , Theorem 2.72 contains Theorem 1.32 as a special case. Since the theory of 1.1, 1.2 and 1.3 could have been developed just as easily for an arbitrary principal ideal ring as scalar domain as for $P[x]$ (except for the statements about finite P -rank), Theorem 1.32 also holds for principal ideal rings. Actually, the only fact essential for the development of 1.1, 1.2 and 1.3 is that every factor ring $\mathfrak{S}/\mathfrak{b}$, where \mathfrak{b} is a non-zero ideal of \mathfrak{S} , is a principal ideal ring. Hence, Theorem 1.32 also holds if \mathfrak{S} consists of the integers of an algebraic number field (see [4], p. 340). Consequently, the criteria for the solvability of systems of linear equations over principal ideal rings and algebraic number fields are also corollaries of Theorem 2.72. Finally, if \mathfrak{S} is a field, the zero-module is an (ω) -prime module and every module M is closed. Hence no a. p.'s different from (ω) occur. Theorem 2.72 says that then a system of linear equations can be solved if and only if the matrix A of the system and the augmented matrix B (see Theorem 2.72) have the same rank. All known criteria of solvability are hence corollaries of Theorem 2.72, and consequently of the criterion of lengths.

2.8. Rank and linear dependence. In 2.7 the notion of rank is defined independently of the notion of linear dependence. In this section we show that for those vector spaces for which it is customary to define the rank in terms of linear dependence, the notions coincide. This fact follows from Lemma 2.81.

LEMMA 2.81. *Let Q be a \mathfrak{p} -prime module of length l of a Noetherian vector space V over \mathfrak{S} as scalar domain. Then, l is the maximum number of linearly independent elements of the factor module V/Q , if $\mathfrak{S}/\mathfrak{p}$ is considered as the scalar domain of V/Q .*

Proof. Since \mathfrak{p} is the fundamental ideal of Q , \mathfrak{p} is the annihilating ideal of V/Q and hence $\mathfrak{s}/\mathfrak{p}$ can be considered as the scalar domain of V/Q . Let $Q \subset Q_1 \subset \cdots \subset Q_{l-1} \subset V = Q_l$ be a \mathfrak{p} -primary composition sequence of Q and let $v_j \in Q_j$ where $v_j \notin Q_{j-1}$ for $1 \leq j \leq l$. Each vector v_j belongs to a coset \bar{v}_j of V/Q and we assert that $\bar{v}_1, \cdots, \bar{v}_l$ are linearly independent with respect to $\mathfrak{s}/\mathfrak{p}$. Suppose that $\Sigma_{j=1}^l \bar{\alpha}_j \bar{v}_j = \bar{0}$ where $\bar{\alpha}_j \in \mathfrak{s}/\mathfrak{p}$ and $\bar{0}$ is the zero element of V/Q . If α_j is a scalar of the coset $\bar{\alpha}_j$ this would imply that $\Sigma_{j=1}^l \alpha_j v_j \in (Q, \mathfrak{p}V)$ and since $\mathfrak{p}V \subseteq Q$, that $\alpha_l v_l \in Q_{l-1}$. However, $v_l \notin Q_{l-1}$ and Q_{l-1} is \mathfrak{p} -prime which implies that $\alpha_l \in \mathfrak{p}$ and hence that $\bar{\alpha}_l = \bar{0}$ where $\bar{0}$ is the zero element of $\mathfrak{s}/\mathfrak{p}$. By repeating this argument for $\Sigma_{j=1}^{l-1} \bar{\alpha}_j \bar{v}_j = \bar{0}$, we find that $\bar{\alpha}_1 = \cdots = \bar{\alpha}_l = \bar{0}$ which proves the assertion. There remains to be proved that, if $\bar{w} \in Q/V$, we can find an $\alpha \neq \bar{w}$ such that $\Sigma_{j=1}^l \bar{\alpha}_j \bar{v}_j = \bar{\alpha} \bar{w}$. Hereto, we assert that $(Q, v_1, \cdots, v_j)(\mathfrak{p}) = Q_j$ for $1 \leq j \leq l$, for $Q \subset (Q, v_1)(\mathfrak{p}) \subseteq Q_1$ and, since $(Q, v_1)(\mathfrak{p})$ is \mathfrak{p} -primary, we conclude that $(Q, v_1)(\mathfrak{p}) = Q_1$ from which the assertion follows by induction. In particular, $(Q, v_1, \cdots, v_l)(\mathfrak{p}) = V$ and hence, if w is a vector in the coset \bar{w} , we can find an $\alpha \notin \mathfrak{p}$ such that $\alpha w = q + \Sigma_{j=1}^l \alpha_j v_j$ where $q \in Q$. This implies that $\bar{\alpha} \bar{w} = \Sigma_{j=1}^l \bar{\alpha}_j \bar{v}_j$ where $\bar{\alpha} \neq \bar{0}$.

The vector spaces for which it is customary to define rank in terms of linear dependence are m -dimensional vector spaces consisting of columns whose m components are elements of a domain of integrity \mathfrak{s} . Hence, for these spaces, the zero-module is a prime module with (ω) as radical. Consequently, the rank of a module M , as defined in Definition 2.72, is then the maximal number of vectors of that module which are linearly independent with respect to \mathfrak{s} , as follows from Lemma 2.81 if we take for Q the (ω) -prime zero-module (0) and as imbedding vector space of (0) the module M .

Remark 2.81. If the scalar domain of a Noetherian vector space has divisors of zero, the notion of linear dependence does not give rise to useful notions of rank. However, Definition 2.72 defines rank with all the desirable useful properties for any scalar domain \mathfrak{s} (with or without divisors of zero) for the case when the zero-module is primary. Moreover, it is never really necessary to make any definition of rank since, according to 2.7, a definition of rank is nothing more than a preferential treatment of that primary component of a module whose a. p. is the radical of the zero-module.

PRINCETON UNIVERSITY,
PRINCETON, NEW JERSEY.

REFERENCES.

- [1] R. A. Frazer, W. J. Duncan and A. R. Collar, *Elementary Matrices*, Cambridge University press (1938).
- [2] A. A. Albert, *Modern Higher Algebra*, University of Chicago Press (1936).
- [3] K. Hentzelt and E. Noether, "Zur Theorie der Polynomideale und Resultanten," *Mathematische Annalen*, vol. 88 (1922), pp. 53-79.
- [4] E. Steinitz, "Rechteckige Systeme und Moduln in algebraischen Zahlkörpern I," *Mathematische Annalen*, vol. 71 (1912), p. 340.
- [5] P. M. Grundy, "A generalization of additive ideal theory," *Proceedings of the Cambridge Philosophical Society*, vol. 38 (1941), pp. 241-279.
- [6] E. Snapper, "Structure of linear sets," *Transactions of the American Mathematical Society*, vol. 52 (1942), pp. 257-264.
- [7] W. Krull, "Idealtheorie," *Ergebnisse der Mathematik*, vol. 4 (1935), pp. 255-402.
- [8] B. L. v. d. Waerden, "Eine Verallgemeinerung des Bezoutschen Theorems," *Mathematische Annalen*, vol. 99 (1928), pp. 497-541.
- [9] B. L. v. d. Waerden, *Moderne Algebra*, vol. 1, Berlin, Springer, 1937.
- [10] B. L. v. d. Waerden, "On Hilbert's function, series of composition of ideals and a generalization of the theorem of Bezout," *Akademie van Wetenschappen te Amsterdam, Proc.*, vol. 31, 1928, pp. 749-770.
- [11] B. L. v. d. Waerden, *Moderne Algebra*, vol. 2, Berlin, Springer, 1940.
- [12] C. Chevalley, "On the theory of local rings," *Annals of Mathematics*, vol. 44 (1943), pp. 690-708.
- [13] E. Snapper, "The resultant of a linear set," *American Journal of Mathematics*, vol. 66 (1944), pp. 59-68.
- [14] C. Chevalley, "On the notion of the ring of quotients of a prime ideal," *Bulletin of the American Mathematical Society*, vol. 50 (1944), pp. 93-97.

SYSTEMS OF TOTAL DIFFERENTIAL EQUATIONS AND LIOUVILLE'S THEOREM ON CONFORMAL MAPPINGS.*

By PHILIP HARTMAN.

1. Let X denote the vector (x_1, x_2, x_3) and let the vector function $X = X(u, v)$, defined on a connected domain R in the (u, v) -plane, be a surface S of class C'' ; so that each component of $X(u, v)$ has continuous partial derivatives of the second order and the vector product

$$(1) \quad (X_u, X_v) \neq 0 \text{ on } R,$$

where the subscript u or v indicates the corresponding partial derivative. Since the surface S is of class C'' , every normal section of S is a curve of class C'' , and so possesses a curvature. At the point $X(u, v)$, let $k_1(u, v)$ and $k_2(u, v)$ denote the principal curvatures, the maximum and minimum curvatures of all normal sections at the point $X(u, v)$. It is known that if every point of S is a "flat point" ("Flachpunkt"), $k_1(u, v) = k_2(u, v) \equiv 0$ on R , then the surface S is part of a plane. Also, if every point of the surface is an "umbilicus" ("Nabelpunkt"), $k_1(u, v) = k_2(u, v) \neq 0$ on R , and if, in addition, it is supposed that $X(u, v)$ is of class C''' , then it is known that the surface S is part of a sphere. Cf., e. g., Bieberbach [2].

It will be shown below that the *additional* assumption of the last statement, requiring that $X(u, v)$ be of class C''' , rather than C'' , is not needed. It can hardly be expected to relax the C'' -condition since the definition of the principal curvatures involves the second fundamental form of the surface. In this direction the following theorem will be proved:

THEOREM 1. *Let $X = X(u, v)$, defined on a connected domain R , be a surface S of class C'' . Let $k_1(u, v)$ and $k_2(u, v)$ denote the principal curvatures and let*

$$(2) \quad k_1(u, v) \equiv k_2(u, v) \text{ on } R.$$

Then the function (2) is constant, and so the surface S is part of either a plane or a sphere, according as this constant is or is not zero.

* Received October 18, 1946.

2. The proof will depend on a theorem on total differentials which may be considered an analogue of the lemma of P. du Bois-Reymond occurring in the calculus of variations. This theorem can be formulated as follows:

THEOREM 2. *Let*

$$(3) \quad x = x(u, v), \quad y = y(u, v)$$

denote scalar functions of class C' on a connected domain U of the (u, v) -plane and suppose that the Jacobian

$$(4) \quad x_u y_v - x_v y_u \neq 0 \text{ on } U.$$

Let $\alpha(u, v)$ be a continuous function on U . A necessary and sufficient condition for

$$(5) \quad \int_C \alpha(u, v) dx(u, v) = 0 \text{ and } \int_C \alpha(u, v) dy(u, v) = 0$$

to hold for every rectifiable Jordan curve C in U is that $\alpha(u, v)$ be a constant.

If the functions $x(u, v)$ and $y(u, v)$ possess continuous partial derivatives of the second order and if the function $\alpha(u, v)$ possesses partial derivatives of the first order, then necessary and sufficient conditions for (5) are

$$(\alpha x_u)_v = (\alpha x_v)_u \text{ and } (\alpha y_u)_v = (\alpha y_v)_u.$$

Performing the partial differentiations indicated in these equations and using the condition (4), one obtains the identities $\alpha_u \equiv \alpha_v \equiv 0$. However, the point in the theorem lies in the fact that these operations are not justifiable under the conditions of Theorem 2, since the existence of the partial derivatives involved is not assumed.

The necessity of the condition (5) is obvious.

Let (u_0, v_0) be an arbitrary point of the domain U . In order to prove the sufficiency of (5), it will be shown that if (5) holds for every rectifiable Jordan curve C in U , then there is a neighborhood of (u_0, v_0) on which $\alpha(u, v)$ is constant.

Let $x_0 = x(u_0, v_0)$ and $y_0 = y(u_0, v_0)$. Then (4) implies that (3) possesses a unique local inverse,

$$(6) \quad u = u(x, y), \quad v = v(x, y),$$

of class C' which transforms some neighborhood,

$$(7) \quad D: |x - x_0| < d, \quad |y - y_0| < d \quad (d > 0),$$

of (x_0, y_0) into a neighborhood of (u_0, v_0) . Thus, it will be sufficient to show that the continuous function

$$(8) \quad \beta(x, y) = \alpha(u(x, y), v(x, y)),$$

defined on (7) , is constant.

Let the curve $C: u = u(t), v = v(t)$, be the image, under the transformation (6), of the rectangle whose vertices are (x_0, y_0) , (x, y_0) , (x, y) , (x_0, y) , where (x, y) is any point of D and the parameter t denotes arc-length on the rectangle (in the (x, y) -plane). Since (6) is of class C' , the Jordan curve C is rectifiable. Thus, the first equation in (5) gives

$$\int_{x_0}^x \beta(t, y_0) dt - \int_{x_0}^x \beta(t, y) dt = 0.$$

Since the function (8) is continuous, it follows from the fundamental theorem of calculus that

$$\beta(x, y_0) = \beta(x, y),$$

if (x, y_0) and (x, y) are points of the domain D . Similarly, the second equation of (5) implies that

$$\beta(x_0, y) = \beta(x, y),$$

if (x_0, y) and (x, y) are points of (7) . The last two formulae mean that $\beta(x, y)$ is constant on (7) . This proves Theorem 2.

3. In order to prove Theorem 1, use will be made of the differential equations of Olinde Rodrigues for the lines of curvature on a surface of class C'' ,

$$(9) \quad N' + \alpha X' = 0,$$

where $N = N(u, v)$ is the unit normal vector, $(X_u, X_v) / |(X_u, X_v)|$; the function $\alpha = \alpha(u, v)$ is either $k_1(u, v)$ or $k_2(u, v)$; and the prime denotes differentiation with respect to a parameter which does not occur explicitly in the differential equations (9).

Under the assumption (2) of Theorem 2, every arc of class C'' on the surface is a line of curvature, that is, a solution of (9). Hence,

$$(10) \quad \int_C \alpha(u, v) dX(u, v) = - \int_C dN(u, v) = 0$$

holds for every Jordan curve C of class C'' on R , the domain of definition of the surface $X(u, v)$. By the 2-dimensional analogue of Helly's term-by-term integration theorems for Stieltjes integrals and in view of standard approximation theorems of the Weierstrass type, (10) holds also for every rectifiable Jordan curve C on R . Since the components of the vector in (1) are the three Jacobians

$$x_{2u}x_{3v} - x_{2v}x_{3u}; \quad x_{3u}x_{1v} - x_{3v}x_{1u}; \quad x_{1u}x_{2v} - x_{1v}x_{2u},$$

it follows from (1) that, at every point of R , at least one of these Jacobians is not zero. Let P be an arbitrary point of R and let x, y denote a pair of the variables x_1, x_2, x_3 selected so that $x_u y_v - x_v y_u \neq 0$ at P . Then, by continuity, (4) holds on some neighborhood U of P . On the other hand, the vector equation (10) implies that (5) holds for every rectifiable Jordan curve C on U . Consequently, by Theorem 2, the continuous function $\alpha(u, v)$ is constant on U . Since the point P of the connected domain R was arbitrary, it follows that the function $\alpha(u, v)$ or (2) is constant on R .

This completes the proof of the first statement of Theorem 1. The second statement is an immediate consequence of the first. For, since $\alpha = \alpha(u, v)$ is a constant, an integration of (9) gives

$$N(u, v) + \alpha X(u, v) = A,$$

a constant vector. If $\alpha = 0$, then the normal vector $N = N(u, v)$ is this constant vector A . But along any arc (of class C') on the surface, the tangent vector is orthogonal to $N = A$. This means that the scalar product $A \cdot X'$ vanishes for all t , so that $A \cdot X = c$, where c is a constant. This shows that the surface S is part of a plane. Similarly, if $\alpha \neq 0$, then

$$X(u, v) = (A - N(u, v))/\alpha,$$

and so, since $N(u, v)$ is a unit vector, the surface S lies on the sphere having the center A/α and radius $|1/\alpha|$.

This completes the proof of Theorem 1.

4. Let

$$(11) \quad Y = Y(X)$$

where $X = (x_1, x_2, x_3)$, $Y = (y_1, y_2, y_3)$ and

$$\det (\partial y_i / \partial x_k) \neq 0$$

be a (schlicht) mapping of class C' of a domain in the X -space upon a domain in the Y -space. The transformation (11) is said to be conformal if there exists a positive, continuous function $\lambda = \lambda(X)$ on T such that

$$dy_1^2 + dy_2^2 + dy_3^2 = \lambda^2(dx_1^2 + dx_2^2 + dx_3^2);$$

with a corresponding definition of conformality in the case of a space with an arbitrary dimension number. A mapping (11) is said to be a Möbius transformation if it can be composed of Euclidean movements, reflections on spheres or planes, and similarity transformations ($Y = cX$, where c is a scalar constant).

Liouville's theorem states that (under conditions of differentiability not specified by him) every conformal mapping is a Möbius transformation if the dimension number of the space exceeds 2. It was pointed out to me by Professor Wintner that the various proofs of Liouville's theorem as given in texts on differential geometry assume either explicitly or implicitly that the conformal mapping (11) is of class C''' . He raised the question whether the class C''' can be reduced to the class C' (just as in the plane case of the Cauchy-Riemann equations, where the Möbius group must be replaced by the group of regular analytic schlicht mappings).

Although the above results do not apply to this extent, they halve the gap between the desideratum and the standard condition, since they reduce the class C''' to the class C'' .

THEOREM 3. *Every 3-dimensional conformal mapping of class C'' is a Möbius transformation.*

A geometrical proof of Liouville's theorem given by Cappelli [4] and reproduced by Blaschke [3] (without bothering with the assumptions to be made) is based on

- (i) Dupin's theorem that if three surfaces of class C'' cut orthogonally, then the arcs of intersection are lines of curvature on each surface, and on
- (ii) the circumstance that, if every direction at every point on the surface is a direction of a line of curvature, then the surface is part of either a plane or a sphere.

Since, as proved above, (ii) holds for surfaces of class C'' , Capelli's proof becomes valid for the sharpened formulation of Liouville's assertion, that is, for Theorem 3.

Remark. Another proof of Liouville's theorem, as given by Bianchi [1], replaces the geometrical considerations (i) and (ii) by Lamé's partial differential equations for orthogonal families. This proof, too, assumes the mapping (11) to be of class C''' . However, due to Theorem 2, the derivation of Lamé's differential equations can be modified so as to lead to another proof of Theorem 3.

The methods of this proof are applicable to the case of any n -dimensional ($n \geq 3$) space. Actually, the previous proof based on (i) and (ii) can be modified to this effect, since Theorems 1 and 2 have n -dimensional analogues.

THE JOHNS HOPKINS UNIVERSITY.

BIBLIOGRAPHY.

- [1] L. Bianchi, *Vorlesungen über Differentialgeometrie*, Leipzig, 1899, pp. 484-488.
- [2] L. Bieberbach, *Differentialgeometrie*, Leipzig and Berlin, 1932, pp. 68-69.
- [3] W. Blaschke, *Vorlesungen über Differentialgeometrie*, vol. I, Berlin, 1930, pp. 100-102.
- [4] A. Capelli, "Sulla limitata possibilità di trasformazioni conformi nello spazio," *Annali di Matematica*, Ser. II, vol. 14 (1886-1887), pp. 227-237.

GENERIC ALGEBRAS.*¹

By W. J. R. CROSBY.

The aim of this paper is to establish a relation between a given algebra class over a suitable type of ground field and certain other algebras, called generic algebras. Suppose that P is the ground field of the given algebra class; then types of central simple algebras are constructed over algebraic function fields with P as field of constants. These algebras are called generic algebras. It is shown that the original algebra class may be obtained from the generic algebra by a specialization of the variables. Thus problems in the theory of algebras may be associated with corresponding problems in the theory of algebraic functions of several variables. In particular it is seen that, in order to consider certain questions on algebras such as that of the possibility of the representation of an algebra as a direct product or crossed product or the existence of cyclic splitting fields of minimal degree, it is sufficient to confine our attention to generic algebras.

After a review in **1** of the theory of algebras needed subsequently, **2** contains a definition of suitable fields of algebraic functions associated with given algebra classes and a definition of the class of related generic algebras. In **3** it is shown how the generic algebra specializes to an algebra of the given algebra class, and in **4** some of the properties of the generic algebras are discussed. Finally, in **5**, it is shown that the properties of a special type of generic algebras, called regular generic algebras, may be more easily discussed.

GENERIC ALGEBRAS.

1. Suppose that S_0 is a central² simple algebra over the field P . Then there exists an algebraic extension F of P , which is normal and separable over P , and which is a splitting field³ for S_0 ; that is

* Received May 18, 1945.

¹ This dissertation was presented to the University of Toronto as part of the requirements for the degree of Doctor of Philosophy. The author wishes to record his gratitude to Professor R. Brauer, under whose supervision the paper was written, and to Dean S. Beatty.

² Or normal.

³ See M. Deuring: *Algebren, Ergebnisse der Mathematik*, vol. 4 (1935), Berlin, p. 48. Also N. Jacobson: *The Theory of Rings*, Mathematical Surveys 2, American Mathematical Society, New York, 1943, p. 104, and B. L. van der Waerden, *Moderne Algebra*, vol. 2 (1940), Berlin, p. 209.

$$S_0 \times F \cong [F]_{r_0},$$

where $[F]_{r_0}$ is the complete matrix algebra of degree r_0 with coefficients in F : The order of S_0 over P is the number r_0^2 . Also, by Wedderburn's theorem,⁴ S_0 is isomorphic to the complete algebra of matrices of some degree t_0 , with coefficients in a division ring D , and this division ring D is unique in the sense of isomorphism; that is

$$S_0 \cong [D]_{t_0} \cong D \times [P]_{t_0}.$$

The division ring D , hereafter referred to as the Wedderburn factor of S_0 , is itself a central (and simple) algebra with center P , and if m^2 is its order over P , then $r_0 = t_0 m$. The unique number m so determined is the index of S_0 .

In the splitting field F there exists an absolutely irreducible representation of the algebra S_0 by means of matrices of order r_0 . The absolutely irreducible representation and its conjugates determine a factor set c of S_0 , the elements of which we shall denote by $c_{A,B,C}$, where A, B, C are elements of the Galois group \mathfrak{G} of F over P . About these numbers $c_{A,B,C}$ we can make the following⁵ statements:

- 1) $c_{A,B,C} \neq 0$, for all A, B, C of \mathfrak{G} .
- 2) $c_{A,B,C}^G = c_{AG,BG,CG}$, where c^G denotes the transform of c by the element G of \mathfrak{G} .
- 3) $c_{A,B,C} \cdot c_{A,C,G} = c_{A,B,G} \cdot c_{B,C,G}$, for all elements A, B, C, G of \mathfrak{G} .
- 4) There is a positive integer e , and there are non-zero numbers $k_{A,B}$ in F , corresponding to each pair of elements A, B of \mathfrak{G} , which form a conjugate double system,⁶ such that

$$c_{A,B,C}^e = \frac{k_{A,B} \cdot k_{B,C}}{k_{A,C}},$$

for all elements A, B, C of \mathfrak{G} . The least positive number e with this property is the exponent of the factor set c in the field F .

- 5) The exponent is a divisor of the index and is divisible by all prime factors of the index.

⁴ See N. Jacobson, *loc. cit.*, p. 98.

⁵ The first three might form a definition of a factor set, the others may be regarded as consequences of the first three and the related theory of algebras. See M. Deuring, *loc. cit.*, p. 52 *et seq.*

⁶ That is, the transform of $k_{A,B}$ by the automorphism G of the Galois group \mathfrak{G} is $k_{AG,BG}$.

6) The factor set c is not uniquely determined by S_0 and F , but may be replaced by the set c' , the elements of which are defined by

$$c'_{A,B,C} = \frac{k_{A,B} \cdot k_{B,C}}{k_{A,C}} \cdot c_{A,B,C},$$

where $k_{A,B}$ is any conjugate double system with no zero elements. The factor set c' is called an associated factor set.

7) If the Wedderburn factor of an algebra S_1 , central over P , is isomorphic to the Wedderburn factor of S_0 , then the algebra S_1 determines, with F , the same class of associated factor sets as that determined by S_0 .

8) The set of all matrices of the form $(z_{A,B} c_{A,B,E})$ where $z_{A,B}$ may be any conjugate double system in F , and E is the unit element of \mathfrak{G} , constitute a central simple algebra of which the order is the square of the order r of \mathfrak{G} , and of which the Wedderburn factor is D , the Wedderburn factor of S_0 .

2. In order to be able to construct suitable algebraic function fields associated with the algebra S , stated to exist in 8), we shall assume⁷ that all the elements of the factor set c lie in P .

Let T denote a $(1-1)$ representation of \mathfrak{G} by means of matrices of degree q with coefficients in P . Such a representation always exists. We shall be concerned with extension fields of P , formed by adjoining q algebraic or transcendental elements to K . If \mathfrak{p} is a column of length q , with elements p_1, p_2, \dots, p_q , we shall write $P(\mathfrak{p})$ for $P(p_1, p_2, \dots, p_q)$; that is, $P(\mathfrak{p})$ is the field of all rational functions, with non-zero denominators, of p_1, p_2, \dots, p_q , with coefficients in P . Similarly, we shall write $f(\mathfrak{p})$ for $f(p_1, p_2, \dots, p_q)$, where f is a rational function of p_1, p_2, \dots, p_q (with non-zero denominator). Let \mathfrak{x} be the column of length q with elements x_1, x_2, \dots, x_q , where x_1, x_2, \dots, x_q are indeterminates over F . Let $\Phi = P(\mathfrak{x})$, and let Π be the field of all elements of Φ which are invariant under the transformations of the representation T of \mathfrak{G} in P . That is, if T_G is the matrix of the representation which corresponds to the element G of \mathfrak{G} , then Π is the field of all rational functions $f(\mathfrak{x})$ of Φ which are such that $f(T_G \mathfrak{x}) = f(\mathfrak{x})$ for every G in \mathfrak{G} . If we write \mathfrak{x}^G for $T_G \mathfrak{x}$, and $f(\mathfrak{x})^G$, or f^G , for $f(T_G \mathfrak{x})$, then we shall have

$$\begin{aligned} (f(\mathfrak{x})^G)^H &= f(T_G \mathfrak{x})^H = f(T_G \mathfrak{x}^H) = f(T_G T_H \mathfrak{x}) \\ &= f(T_{GH} \mathfrak{x}) = f(\mathfrak{x}^{GH}) = f^{GH}, \end{aligned}$$

as we should expect.

⁷ We shall see later that we may instead assume that P contains certain roots of unity.

The field Φ is normal⁸ over Π . For, if f is any element of Φ , and t is an indeterminate, then evidently the function $\chi(t) = \prod_G (t - f^G)$, where the product is taken over all elements G of \mathfrak{G} , is a polynomial with coefficients in Π . All the roots of $\chi(t) = 0$ lie in Φ , and one of the roots is f . Since every element of Φ is the root of an equation of degree less than or equal to r with coefficients in Π , Φ is normal over Π with degree not greater than r . But certainly all the elements of \mathfrak{G} define automorphisms of Φ over Π . Hence the degree of Φ over Π is exactly r . The Galois group of Φ over Π is evidently \mathfrak{G} . Also the element G of \mathfrak{G} carries f into f^G , in accordance with the notation used above.

We next define an algebra Σ by means of Φ in the same way that we defined S by means of F . Since the factor set c lies in P , it must lie in Φ . Let $\xi_{A,B}$ be a conjugate double system of elements of Φ . Let Σ be the set of all matrices of the form $(\xi_{A,B} c_{A,B,E})$. Then Σ will be called the generic algebra associated with S ; it is a central simple algebra of order r with center Π , and is determined by the fields P and F and the factor set c . In order to learn in what ways Σ over Π resembles S over P , we try to replace the indeterminates of Φ by elements of F , in such a way that, in this process of specialization, Φ becomes F , Π becomes P , and Σ becomes S .

3. The results on specialization of variables which will be established all depend on the properties of a certain matrix. In our first lemma we prove the existence of this matrix. The letter \mathfrak{p} , in all our arguments, will be used to denote a column of length q with elements p_1, p_2, \dots, p_q ; the numbers p_1, p_2, \dots, p_q will always be assumed to be elements of P .

LEMMA 1. *There exists a matrix R of degree q with coefficients in F , such that*

$$(i) \quad R^G = T_G R;$$

$$(ii) \quad R \text{ is non-singular};$$

(iii) *If the vector \mathfrak{p} is invariant under no transformation T_G except when G is the unit element E of \mathfrak{G} , then $R\mathfrak{p}$ is invariant for only the unit element E of \mathfrak{G} ; that is, if the relation $T_G \mathfrak{p} = \mathfrak{p}$ implies that $G = E$, then the relation $(R\mathfrak{p})^G = R\mathfrak{p}$ implies that $G = E$;*

(iv) *If the representation T is irreducible, and \mathfrak{p} is not zero, then $R\mathfrak{p}$ is invariant for only the unit element E of \mathfrak{G} .*

⁸ Cf., for instance, E. Artin, *Galois Theory*, Notre Dame Mathematical Lectures 2, Notre Dame, 1942, p. 31.

Proof. Let v be a number of F such that the transforms of v by the elements of \mathfrak{G} are linearly independent⁹ with respect to P . Then the system of numbers $u_G = v^{G^{-1}}$ forms a basis for F over P , and

$$u_G^H = v^{G^{-1}H} = v^{(H^{-1}G)^{-1}} = u_{H^{-1}G}$$

for every pair of elements G, H of \mathfrak{G} . We now prove that we may take

$$R = \sum_G u_G T_G,$$

where the summation extends over all elements of \mathfrak{G} .

The statement (i) is certainly true, for, writing L for $G^{-1}H$, we have

$$\begin{aligned} R^G &= \sum_H u_H^G T_H = \sum_H u_{H^{-1}H} T_H \\ &= \sum_L u_L T_{GL} = T_G \sum_L u_L T_L = T_G R. \end{aligned}$$

Next let η be a column of length q with elements y_1, y_2, \dots, y_q , and let us suppose that the equation $R\eta = 0$ has a non-trivial solution. Let $F = P(\theta)$ where θ is an element of F and let us write θ_G for θ^G . Then since $R\eta = 0$, $(R\eta)^G = 0$. Therefore $R(\eta^G) = 0$, since $R^G = T_G R$ and T_G is non-singular. Consequently

$$R \sum_G \theta_G^j \eta^G = \sum_G \theta_G^j R(\eta^G) = 0$$

for any positive integer j . But $\sum_G \theta_G^j \eta^G$ is different from zero if j is suitably chosen and all its elements lie in P . We have therefore proved that if the equation $R\eta = 0$ has a non-trivial solution in F , it has a non-trivial solution in P . But if we now suppose that $R\eta = 0$ where y_1, y_2, \dots, y_q lie in P , then from the definition of R , $\sum_G u_G T_G \eta = 0$, and consequently $T_G \eta = 0$ for every G in \mathfrak{G} , since the system of numbers u_G forms a basis for F over P , and the elements of T_G are in P . Therefore $\eta = 0$, since T_G is non-singular. The equation $R\eta = 0$ has, therefore, only the trivial solution $\eta = 0$. Consequently R is non-singular. This is part (ii) of the lemma.

Let us now suppose that $(R\eta)^G = R\eta$. Employing part (i) of the lemma, we can write this relation in the form $(T_G - I)R\eta = 0$, where I is the unit matrix, and substituting our expression for R , we find

$$\sum_H (T_G - I) u_H T_H \eta = 0.$$

From this equation we conclude that

$$(T_G - I) T_H \eta = 0$$

for every H in \mathfrak{G} . This implies that $T_{GH}\eta = T_H\eta$ and hence, that

⁹ See E. Artin, *loc. cit.*, p. 53.

$$(1) \quad T_{H^{-1}GH}p = p$$

for every H in \mathfrak{G} .

The statement (iii) now follows; for if the relation $T_G p = p$ implies that $G = E$, the equation (1) implies that $H^{-1}GH = E$ and therefore $G = E$.

Finally, to prove (iv), we first note that, if $(Rp)^G = Rp$, then from the equation (1) which we have established we conclude that p is invariant under any transformation T_M , if M is any conjugate element or product of conjugate elements of G . Let \mathfrak{M} be the class of all the conjugates and products of conjugates of G . We easily verify that \mathfrak{M} is an invariant subgroup of \mathfrak{G} . Assuming that p is not zero, let us take the elements of p to be the components of the first of q linearly independent vectors in a q -dimensional vector space over P . Then the set of vectors with only the first component different from zero constitutes an invariant subspace with respect to the linear transformations defined by the representation T_M of the group \mathfrak{M} . Therefore, there must be a non-singular matrix Q with coefficients in P , such that

$$Q^{-1}T_M Q = \begin{pmatrix} 1 & 0 \\ 0 & U_M \end{pmatrix}$$

for every M in \mathfrak{M} , where U_M is a representation of \mathfrak{M} of degree $q - 1$. Therefore, by Clifford's Theorem,¹⁰ the representation T_M of \mathfrak{M} reduces completely to the unit matrix. Consequently the subgroup \mathfrak{M} must contain only the unit element of \mathfrak{G} . Thus $(Rp)^G = Rp$ implies that G is the unit element of \mathfrak{G} .

We can now discuss the result of replacing the indeterminates in $\Phi = P(x)$ by suitably chosen elements of F . It is, of course, to be understood that elements of which the denominators are made to vanish by the substitution are to be discarded.

THEOREM 1. *If we replace x in $\Phi = P(x)$ by Rp , the resulting specialized field, namely $P(Rp)$, coincides with F , provided either that p satisfies no equation of the form $T_G p = p$ except when G is the unit element E of \mathfrak{G} , or that T is irreducible and p is not zero.*

Proof. Since $P(Rp)$ contains P and is contained in F , $P(Rp)$ must be the field of invariants for some subgroup \mathfrak{H} of the Galois group \mathfrak{G} of F over P . Then Rp must be invariant under the transformations of \mathfrak{H} . That is, $(Rp)^G = Rp$ if G is in \mathfrak{H} . Therefore, by Lemma 1, \mathfrak{H} contains only the unit element of \mathfrak{G} . Therefore $P(Rp) = F$.

¹⁰ See H. Weyl, *The Classical Groups*, Princeton, 1939, p. 159.

THEOREM 2. *If we replace \mathfrak{x} in each function $f(\mathfrak{x})$ of Π by $R\mathfrak{p}$, the resulting specialized field coincides with P .*

Proof. Let us denote the specialized field by P_1 . Evidently F contains P_1 . Since Π contains P , P_1 must contain P . Now Π was defined to be the set of all rational functions $f(\mathfrak{x})$, with coefficients in P , such that $f(T_G \mathfrak{x}) = f(\mathfrak{x})$. If we replace \mathfrak{x} by $R\mathfrak{p}$ in these functions, and discard those expressions of which the denominator vanishes, we find that, for every element $f(R\mathfrak{p})$ of P_1 ,

$$f(R\mathfrak{p})^G = f(R^G \mathfrak{p}) = f(T_G R\mathfrak{p}) = f(R\mathfrak{p})$$

for every G in \mathfrak{G} . It follows that P_1 is contained in P , and therefore $P_1 = P$.

THEOREM 3. *If we replace \mathfrak{x} by $R\mathfrak{p}$ in Σ , the specialized algebra coincides with S , provided either that \mathfrak{p} satisfies no equation of the form $T_G \mathfrak{p} = \mathfrak{p}$ except when G is the unit element E of \mathfrak{G} , or that T is irreducible and \mathfrak{p} is different from zero.*

Remark. It is to be understood that, if the substitution of $R\mathfrak{p}$ for \mathfrak{x} makes the denominator of one of the coefficients of a matrix of Σ vanish, that matrix is to be discarded.

Proof. It is evidently sufficient to prove that (i) if $\zeta_{A,B}$ is a conjugate double system of elements of Φ , and if the system $\zeta_{A,B}$ specializes to the system $z_{A,B}$ in F , when we replace \mathfrak{x} by $R\mathfrak{p}$, then $z_{A,B}$ is a conjugate double system in F , and also that (ii) if $z_{A,B}$ is any conjugate double system of elements of F , there exists in Φ a conjugate double system $\zeta_{A,B}$ which reduces to $z_{A,B}$ when we replace \mathfrak{x} by $R\mathfrak{p}$. The first statement is certainly true, for if $\zeta_{A,B} = \zeta_{A,B}(\mathfrak{x})$, then $z_{A,B} = \zeta_{A,B}(R\mathfrak{p})$, and consequently

$$\begin{aligned} z_{A,B}^G &= \zeta_{A,B}(R\mathfrak{p})^G = \zeta_{A,B}(T_G R\mathfrak{p}) \\ &= \zeta_{AG,BG}(R\mathfrak{p}) = z_{AG,BG}. \end{aligned}$$

In order to establish (ii), let us suppose that $z_{A,B}$ is a conjugate double system in F . Let E be the unit element of \mathfrak{G} . Then by Theorem 1, there are elements $\zeta_{E,A} = \zeta_{E,A}(\mathfrak{x})$ in Φ , which specialize to the elements $z_{E,A}$ in F for each A in \mathfrak{G} . We then define $\zeta_{A,B}$ to be $\zeta_{E,BA^{-1}}^A = \zeta_{E,BA^{-1}}(T_A \mathfrak{x})$. The system so defined is a conjugate double system in Φ , since

$$\begin{aligned} \zeta_{A,B}^G &= \zeta_{E,BA^{-1}}^A(T_A \mathfrak{x}^G) = \zeta_{E,BA^{-1}}(T_A T_G \mathfrak{x}) \\ &= \zeta_{E,(BG)(AG)^{-1}}(T_{AG} \mathfrak{x}) = \zeta_{AG,BG}. \end{aligned}$$

Also, $\zeta_{A,B}$ specializes to $z_{A,B}$ for every A, B of \mathfrak{G} , since

$$\begin{aligned}\xi_{A,B}(Rp) &= \xi_{E,BA^{-1}}(T_A Rp) = \xi_{E,BA^{-1}}(R^A p) \\ (\xi_{E,BA^{-1}}(Rp))^A &= z_{E,BA^{-1}}^A = z_{A,B}.\end{aligned}$$

The proof is now complete.

4. In order to discover the properties of Σ it is necessary to examine a little more closely the conjugate double systems $\xi_{A,B}$, which are characterized by the relation $\xi_{A,B}^G = \xi_{AG,BG}$. Suppose $\xi_{A,B} = \xi_{A,B}(x)/\eta_{A,B}(x)$, where $\xi_{A,B}(x)$ and $\eta_{A,B}(x)$ are polynomials which are prime to each other. Then since

$$\left(\frac{\xi_{A,B}(x)}{\eta_{A,B}(x)}\right)^G = \frac{\xi_{A,B}(T_G x)}{\eta_{A,B}(T_G x)} = \frac{\xi_{AG,BG}(x)}{\eta_{AG,BG}(x)},$$

we have

$$(1) \quad \xi_{A,B}(T_G x) \cdot \eta_{AG,BG}(x) = \xi_{AG,BG}(x) \cdot \eta_{A,B}(T_G x).$$

Therefore $\xi_{AG,BG}(x)$, being prime to $\eta_{AG,BG}(x)$, must divide $\xi_{A,B}(T_G x)$. Consequently

$$\xi_{A,B}^G(x) = \xi_{A,B}(T_G x) = m_{A,B,G} \xi_{AG,BG}(x),$$

where $m_{A,B,G}$ is an element of P , since $\xi_{A,B}(T_G x)$ and $\xi_{AG,BG}(x)$ have the same degree. From the relation (1) we see that we have also

$$\eta_{A,B}^G(x) = \eta_{A,B}(T_G x) = m_{A,B,G} \eta_{AG,BG}(x).$$

Let $m_G = (\Pi_A \Pi_B m_{A,B,G})^2$, where A and B run through the elements of \mathfrak{G} . Similarly let $M(x) = \Pi_A \Pi_B \xi_{A,B}(x) \cdot \eta_{A,B}(x)$. Then

$$M^G = M(T_G x) = m_G M(x).$$

We next prove a result which is fundamental in all our subsequent discussions of specializations of Φ .

LEMMA 2. *If $M(x)$ is any non-zero polynomial element of Φ such that $M^G = M(T_G x) = m_G M(x)$ for every G in \mathfrak{G} , where m_G is an element of P , then there is some p , not satisfying any equation of the form $T_G p = p$ except when G is the unit element E of \mathfrak{G} , for which $M(Rp)$ is different from zero.*

Proof. Let us suppose that $M(Rp) = 0$ for every p which does not satisfy an equation of the form stated. Let us as before assume that $F = P(\theta)$, where θ is an element of F , and let us write θ_G for θ^G . Then if

$$H(p) = \Sigma_G (\theta^G M(Rp))^G,$$

where G runs through the elements of \mathfrak{G} , and j is a positive integer to be suitably chosen, then

$$H(p) = \sum_G \theta_G^j M^G(Rp) = \sum_G \theta_G^j M(T_G Rp) = (\sum_G \theta_G^j m_G) M(Rp).$$

The coefficient of $M(Rp)$ on the right side is different from zero if j is suitably chosen. The expression $H(p)$ has elements in P which vanish for every p which does not satisfy any equation of the form stated. If we observe that $\Pi'_G((T_G - I)p)H(p)$, where G runs through all elements except the unit element of \mathfrak{G} , is a matrix which vanishes for every p , we easily conclude that $H(p)$ vanishes for every p . Therefore the coefficients of p_1, p_2, \dots, p_q in $H(p)$ must vanish, and consequently the coefficients of p_1, p_2, \dots, p_q in $M(Rp)$ must vanish. It follows that we may replace p in $M(Rp)$ by any matrix of order q with zeros in the 2-nd, 3-rd, \dots , q -th columns, and the resulting expression must vanish. We therefore replace p by $R^{-1}x$, which exists since R is non-singular (by Lemma 1), and we conclude that $M(x) = 0$ for every x . But this is a contradiction. Therefore the lemma is true.

From our preliminary study of conjugate double systems we obtain

COROLLARY 1. *If $\zeta_{A,B}$ is a conjugate double system in Φ , then there exists a specialization of Φ into F such that the system $\zeta_{A,B}$ in Φ becomes a well defined non-zero conjugate double system $z_{A,B}$ in F .*

COROLLARY 2. *If $f(x)$ is a non-zero element of Π , then there exists a specialization of Φ into F such that $f(x)$ becomes a well defined non-zero element of P .*

For we need only take M in the lemma¹¹ to be the product of the numerator and denominator of $f(x)$. The hypotheses of the lemma are easily seen to be satisfied, by an argument analogous to that used in the discussion of conjugate double systems.

From the first corollary we obtain

COROLLARY 3. *Given any non-zero element σ of Σ , there is a specialization of Φ into F such that σ becomes a well defined non-zero element s of S .*

Finally, we also have

COROLLARY 4. *Given any finite number of conjugate double systems in Φ , any finite number of non-zero elements of Π , and any finite number of non-zero elements of Σ , there exists a specialization of Φ into F such that*

¹¹ Or we may regard Corollary 2 as a consequence of Corollary 1.

these become well defined conjugate double systems in F , non-zero elements of P , and non-zero elements of S respectively.

For we can take M in the lemma to be the product of all the non-zero numerators and denominators of all the elements of Φ which enter into the definition of these expressions.

We can now consider some of the properties of the algebra Σ .

THEOREM 4. *If ϵ is the exponent of the factor set c in the field Φ , and e is the exponent in the field F , then e is a divisor of ϵ .*

Proof. The number ϵ is the smallest positive integer such that there exists a conjugate double system $\zeta_{A,B}$ in Φ , such that

$$c^{\epsilon}_{A,B,C} = \frac{\zeta_{A,B}\zeta_{B,C}}{\zeta_{A,C}}.$$

Let us now, using Corollary 1 of Lemma 2, choose a specialization of Φ into F by which the system $\zeta_{A,B}$ becomes a conjugate double system $z_{A,B}$ in F . Then

$$c^e_{A,B,C} = \frac{z_{A,B}z_{B,C}}{z_{A,C}}.$$

Therefore e divides ϵ , since e is the minimum exponent with this property.

THEOREM 5. *The index m of S is a divisor of the index μ of Σ .*

Proof. Suppose that D, Δ are the Wedderburn factors of S, Σ respectively, so that

$$S \cong D \times [P]_t, \text{ and } \Sigma \cong \Delta \times [\Pi]_\tau,$$

where $mt = r = \mu\tau$. We may write $\Delta = \delta\Sigma\delta$, where δ is a primitive idempotent element of Σ . Let us now choose a specialization of Φ to F such that δ becomes a well defined element d of S . Then Δ becomes an algebra $D_1 = dSd$ over P . Also, Σ becomes S , and $[\Pi]_\tau$ becomes $[P]_\tau$. Therefore

$$S \cong D_1 \times [P]_\tau.$$

Since $[P]_\tau$ and S are central and simple, D_1 also is central and simple.¹² If now the Wedderburn factor of D_1 is D_2 , then $D_1 \cong D_2 \times [P]_u$, where u is some positive integer, and consequently

$$S \cong D_1 \times [P]_\tau \cong D_2 \times [P]_u \times [P]_\tau \cong D_2 \times [P]_{u\tau}.$$

¹² See A. A. Albert, "On direct products, cyclic division algebras and pure Riemann matrices," *Transactions of the American Mathematical Society*, vol. 33 (1931), p. 221.

It follows then, from Wedderburn's theorem, that $D_2 \cong D$ and $u\tau = t$. From the relation $mt = \mu\tau$, we find that $\mu = um$. Therefore m is a divisor of μ .

THEOREM 6. *If Σ has a normal splitting field Ω of degree s over Π , then S has a normal splitting field V of which the degree over P is a divisor of s . The Galois group of V over P is (isomorphic to) a sub-group of the Galois group of Ω over Π .*

Proof. Let \mathfrak{B} be the Galois group of Ω over Π , and let $\Omega = \Pi(\omega)$, where ω is a root of the separable and irreducible equation

$$\phi(t) = t^s + \alpha_1 t^{s-1} + \cdots + \alpha_s = 0.$$

The coefficients $\alpha_1, \alpha_2, \cdots, \alpha_s$ are elements of Π . Let δ be the discriminant of $\phi(t)$, and let $\omega_E = \omega, \omega_A, \omega_B, \cdots$ be the conjugates of ω , where E is the unit element, and A, B, \cdots are the other elements of the Galois group \mathfrak{B} . Then there are relations of the form

$$\omega_A = \gamma_A(\omega) \text{ and } \gamma_A(\gamma_B(\omega)) = \gamma_{AB}(\omega)$$

for every A, B of \mathfrak{B} , where $\gamma_A(t)$ is a polynomial of degree less than s , with coefficients in Π .

Let $\lambda_1(t), \lambda_2(t), \cdots, \lambda_s(t)$ be the elementary symmetric functions of degree $1, 2, \cdots, s$ respectively in the s functions of the form $\gamma_A(t)$ for A in \mathfrak{B} . Then

$$\lambda_i(\omega) = (-1)^i \alpha_i \quad (i = 1, 2, \cdots, s).$$

Since $\phi(t)$ is irreducible, there exist polynomials $v_{A,B}(t)$ for every pair of elements A, B of \mathfrak{B} , and polynomials $\kappa_i(t)$ for $i = 1, 2, \cdots, s$, such that

$$\gamma_A(\gamma_B(t)) = \gamma_{AB}(t) + v_{A,B}(t)\phi(t)$$

and

$$\lambda_i(t) = (-1)^i \alpha_i + \kappa_i(t)\phi(t).$$

Let us now choose a specialization of Φ into F such that δ and all the coefficients of all the polynomials $\phi, \gamma_A, v_{A,B}$ and κ_i become non-zero elements of P . Then if $\phi(t)$ becomes $f(t) = t^s + a_1 t^{s-1} + \cdots + a_s$, and $\gamma_A(t), v_{A,B}(t), \lambda_i(t), \kappa_i(t)$ become $g_A(t), n_{A,B}(t), l_i(t), k_i(t)$ respectively, we have

$$g_A(g_B(t)) = g_{AB}(t) + n_{A,B}(t)f(t) \quad (A, B \text{ in } \mathfrak{B}),$$

and

$$l_i(t) = (-1)^i a_i + k_i(t)f(t) \quad (i = 1, 2, \cdots, s).$$

Let w be a root of the equation $f(t) = 0$, and let $V = P(w)$. Further, let

$\psi(t)$ be an irreducible factor of $f(t)$ such that $\psi(w) = 0$. Then the above relations show that each root of the equation $\psi(t) = 0$ can be written as $g_A(w)$, where A is an element of a subset \mathfrak{B} of \mathfrak{A} . Since δ becomes a non-zero element of P , $f(t)$ and therefore also $\psi(t)$ must be separable. We have thus seen that $\psi(t)$ is normal, separable, and irreducible. Automorphisms of V over P may be defined by the transformation definition $w^A = g_A(w)$, for A in \mathfrak{B} , since $g_A(g_B(w)) = g_{AB}(w)$. Evidently \mathfrak{B} is a subgroup of \mathfrak{A} , and V is normal over P with Galois group \mathfrak{B} . Now Ω may be considered to be the field of all polynomials, modulo $\phi(t)$, with coefficients in Π . Suppose $\beta(t)$ is any polynomial with coefficients in Π , which specializes to a polynomial $b(t)$ with coefficients in P . Then division by $\phi(t)$ gives a relation

$$\beta(t) = \phi(t)\pi(t) + \rho(t),$$

where $\tau(t)$ and $\rho(t)$ are polynomials with coefficients which are integral rational functions of the coefficients of $\beta(t)$ and $\phi(t)$, and where the degree of $\rho(t)$ is less than s . Therefore $\pi(t)$ and $\rho(t)$ must specialize to polynomials $p(t)$ and $r(t)$ respectively, and we have

$$b(t) = f(t)p(t) + r(t).$$

Consequently Ω specializes to a commutative algebra W consisting of all polynomials, modulo $f(t)$, with coefficients in P . Since $f(t)$ is in general reducible, W is in general not a field. The field Ω being a splitting field for Σ , there must exist in Ω an absolutely irreducible representation Λ of degree r . In this representation the element κ of Π may be assumed to be mapped on κI . For, if Q is the image of κ , and if k is a characteristic root of the matrix Q , then $\Lambda(Q - kI) = (Q - kI)\Lambda$, and consequently, by Schur's lemma, $Q - kI = 0$, that is, $Q = kI$. If $\kappa \neq k$, the correspondence $k \rightarrow \kappa$ defines an automorphism of P . By applying to Λ the inverse automorphism we obtain an absolutely irreducible representation in which the image of κ is κI . Let $\sigma_1, \sigma_2, \dots, \sigma_{r^2}$ form a basis of Σ over Π , let $\Lambda_1, \Lambda_2, \dots, \Lambda_{r^2}$ be their representatives in Λ , and let $\Lambda_v = (\lambda_{i,j}^{(v)})$ for $v = 1, 2, \dots, r$. Then $\lambda_{i,j}^{(v)}$, being an element of Ω , may be considered to be a polynomial of degree less than s , with coefficients in Π . Let us choose a specialization of Φ into F in such a way that the requirements of the first part of our argument are satisfied, and also in such a way that all the coefficients in the polynomials $\lambda_{i,j}^{(v)}$ become non-zero elements of P . Then $\Lambda_1, \Lambda_2, \dots, \Lambda_{r^2}$ become matrices with coefficients in W , which define a representation of S . But since we may obtain V from W by identifying elements of W of which the difference is divisible by $\psi(t)$, the representation of S in

W determines a representation of S in V . Since this representation in V has degree r , it must be absolutely irreducible, and therefore V must be a splitting field of S .

Since a subgroup of a cyclic group is cyclic, we obtain the

COROLLARY. If Σ has a cyclic splitting field of degree s over Π , then S has a cyclic splitting field of which the degree over S is a divisor of s .

5. The results of the previous sections have been derived without making any special assumptions about the splitting field F and the factor set c of S . In order to obtain more precise results we now assume that the splitting field F is regular;¹³ that is, we assume that F has been constructed in such a way that the elements of the factor set c are all roots of unity.

Suppose that $F_0 = P(\eta)$ is any separable splitting field of S of degree v over P . Let F_1 be the normal field obtained by adjoining all the conjugates of η to P , and let \mathfrak{G}_1 be the Galois group of F_1 over P . In F_1 we can construct a factor set c_1 with exponent e_1 . If now we adjoin to F_1 the n -th roots of suitable elements of F_1 , where n is a multiple of e_1 , we obtain a regular field F . In F the factor set c_1 is associated with a factor set c of which all the elements are n -th roots of unity. (If \mathfrak{G} is the Galois group of F over P , then the definition of c_1 is first extended to the group \mathfrak{G} by the natural homomorphism of \mathfrak{G} on \mathfrak{G}_1 .) Let us now denote by \mathfrak{N} the group of n -th roots of unity in F . Then the rules

$$Z_A Z_B = Z_{ABC}^{-1}{}_{AB,B,E}$$

$$\lambda Z_A = Z_A \lambda^A,$$

where E is the unit element, and A, B are any elements of \mathfrak{G} , and λ is in \mathfrak{N} , define a multiplicative group \mathfrak{S} consisting of all products of the form λZ_A . The associativity of multiplication is secured by the property (3) of factor sets which was mentioned in Section 1. The group \mathfrak{N} is an invariant subgroup of \mathfrak{S} , and $\mathfrak{S}/\mathfrak{N} \cong \mathfrak{G}$.

The assumption made at the beginning of Section 2, namely, that P contains the elements of the factor set c , is now satisfied if we make the weaker and at the same time more simple assumption that P contains the n -th roots of unity.

¹³ For the details of this theory, of which only an outline is given here, see R. Brauer, "Über die Konstruktion der Schiefkörper, die von endlichem Rang in bezug auf ein gegebenes Zentrum sind," *Journal für die Reine und Angewandte Mathematik*, vol. 168 (1932), pp. 59-60.

In terms of the group \mathfrak{G} , defined by the regular splitting field F , we now state¹⁴ a theorem due to R. Brauer by which the exponent ϵ is determined explicitly.

THEOREM 7. *If the factor set c lies in P , and if \mathfrak{G} is the group associated with c , then the exponent with respect to Φ and Π of the factor set c is the smallest positive number ϵ for which there is a subgroup \mathfrak{K} of \mathfrak{G} with the following properties: 1) \mathfrak{K} contains the commutator subgroup of \mathfrak{G} . 2) \mathfrak{K} has an intersection of order ϵ with the group \mathfrak{N} of n -th roots of unity. 3) There is a number t , such that \mathfrak{K} contains the t -th powers of all group elements of \mathfrak{G} and at the same time P contains the t -th roots of unity.*

We next employ the regular field F and the group \mathfrak{G} associated with c to obtain a more precise result about the index μ .

Let \mathfrak{S} be the subgroup of \mathfrak{G} which belongs to the subfield F_0 of F . Then¹⁵ $c_{A,B,C} = 1$ if A, B, C are in \mathfrak{S} , and consequently the elements Z_S of \mathfrak{G} with S in \mathfrak{S} form a subgroup \mathfrak{N} of \mathfrak{G} which has only the unit element of \mathfrak{G} in common with \mathfrak{N} . The degree of F_0 over P being v , the order of \mathfrak{S} , and therefore also of \mathfrak{N} must be r/v . According to an established property¹⁶ of the group \mathfrak{G} , the index of any algebra possessing the factor set with which \mathfrak{G} is associated is a divisor of $r/(r/v)$, i. e. of v . Since S has a splitting field of order equal to its index m , we may take $v = m$, and since the algebra \mathfrak{Z} possesses the same factor set as does A , the above argument shows that the index μ of \mathfrak{Z} is a divisor of m . Combining this result with Theorem 5, we see that we have thus proved

THEOREM 8. *If the regular splitting field F is constructed by extending a splitting field of S of degree m over P in the way described, then $\mu = m$.*

An algebra which can be defined, as we have defined S over P and \mathfrak{Z} over Π , by means of a normal field and a factor set contained in the field, is called a crossed product of the normal field and its Galois group. Let us suppose that S has a normal splitting field V of degree m over P . Then we may define a factor set of S in V , and with this factor set and the field V we may define a crossed product algebra B . Since the order of B is m^2 , B must be isomorphic to D , the Wedderburn factor of both S and B . That is, if A has a normal splitting field of degree m over P , then D is a crossed

¹⁴ Without proof. See R. Brauer, "Über die Kleinsche Theorie der Algebraischen Gleichungen," *Mathematische Annalen*, vol. 110 (1934), p. 500.

¹⁵ See R. Brauer, *loc. cit.*¹³, p. 60.

¹⁶ See R. Brauer, *loc. cit.*¹³, p. 58.

product. Similarly, if Σ has a normal splitting field of degree μ over Π , then Δ is a crossed product. From Theorems 6 and 8 we therefore obtain

THEOREM 9. *If the hypotheses of Theorem 8 are true, and if Δ is a crossed product, then D is a crossed product.*

Crossed products with a cyclic field, that is, cyclic algebras, are of particular interest. From the corollary to Theorem 6 we derive the

COROLLARY. *If the hypotheses of Theorem 8 are true, and if Δ is a cyclic algebra, then D is a cyclic algebra.*

From Theorem 9 we obtain directly ¹⁷

THEOREM 10. *If there exist central division algebras which are not crossed products, then there exist generic algebras which are not crossed products.*

THEOREM 11. *If the hypotheses of Theorem 8 are true, and if Δ is a direct product, then D is a direct product.*

Proof. We have seen, in the proof of Theorem 5, that Δ specializes to an algebra $D_1 = D \times [P]_u$, where $\mu = um$. By Theorem 8, $u = 1$ and consequently $D_1 = D$. Suppose that $\Delta = \Delta_2 \times \Delta_3$, where Δ_2, Δ_3 are subalgebras of Δ . Let $\eta_1, \eta_2, \dots, \eta_h$ be a basis for Δ_2 over Π , and let $\zeta_1, \zeta_2, \dots, \zeta_k$ be a basis for Δ_3 over Π . Each η commutes with each ζ . We now choose a specialization such that the elements $\eta_1, \eta_2, \dots, \eta_h$ and the elements $\zeta_1, \zeta_2, \dots, \zeta_k$, which are in Δ , and therefore also in Σ , become well defined systems y and z in S . Evidently each y commutes with each z . Therefore Δ specializes to $D = D_2 \times D_3$, where D_2 and D_3 are central simple ¹⁸ subalgebras of D .

UNIVERSITY OF SASKATCHEWAN,
SASKATOON, CANADA.

¹⁷ It is not yet known whether there exist central division algebras which are not crossed products. This theorem might afford a means of discussing this problem.

¹⁸ See A. A. Albert, *loc. cit.*

SIMPLE LINKS AND FIXED SETS UNDER CONTINUOUS MAPPINGS.*

By J. L. KELLEY.

In generalizing a result of W. L. Ayres¹ I have shown² that a homeomorphism of any compact continuum into itself leaves invariant a certain sub-continuum which is without cut points. From this conclusion it follows that under a homeomorphism either there is a fixed point or some 0-th order cyclic element is carried into a subset of itself (according to G. T. Whyburn, a 0-th order cyclic element is a continuum maximal with respect to the property of containing no cut points of itself³). In attempting to obtain related results for any continuous transformation R. L. Moore's decomposition into *simple links* is found to be more suitable.⁴ In what follows new results on simple links are proven, and in logically developing the theory some of Moore's results are repeated.

The last section is devoted to transformation theory. It is shown that for any transformation T of a compact continuum M into itself, there exists a continuum π , a subset of a simple link of M , such that $T(\pi) \supset \pi$. Corollaries to this theorem include the theorem of Ayres previously men-

* Received October 2, 1946; Presented to the Society December, 1938. This manuscript was offered to *Fundamenta Mathematicae* in June, 1939. An abstract appeared in *Proceedings of the National Academy of Sciences*, vol. 26 (1940), pp. 192-194. Two changes in the nomenclature of the abstract have been made, namely, instead of F -set, simple link; instead of B -set, central set.

Since this paper was written a unified treatment of cyclic element theory has appeared in G. T. Whyburn's *Analytic Topology*, American Mathematical Society Colloquium Publications. Certain of the results of the first sections of the paper appear in this book.

The author has benefited considerably from discussions of these topics with Professor Whyburn.

¹ See *Fundamenta Mathematicae*, vol. 16 (1930), pp. 333-336.

² See *Duke Mathematical Journal*, vol. 5 (1939), pp. 535-537.

³ See *Bulletin of the American Mathematical Society*, vol. 40 (1934), pp. 159-165.

⁴ See *Foundations of point set theory*, American Mathematical Colloquium Publications, p. 72. The definition of simple link given here differs from Moore's only in that all cut points are simple links. This change is made to conform to cyclic element theory for Peano curves.

The first half of 2.1 and Theorems 2.4 and 2.5 of this paper are contained in Moore's results. The methods of proof differ considerably from Moore's.

tioned and the Scherrer theorem⁶ that any dendrite has the fixed point property.

1. Preliminaries. Throughout, M will denote a compact metric continuum.

*Definition.*⁷ A continuum $A \subset M$ is a *nodal set* if $A \cap \overline{M - A}$ consists of at most a single point. A continuum B is a *central set* if $\overline{M - B}$ consists of a finite number of components, each cutting B in a single point. Note that M itself is a central set.

We shall need the following properties of nodal sets:

- 1.1. a) If A is a nodal set so is $\overline{M - A}$.
- b) If A is a nodal set and C is connected, then $A \cap C$ is connected or vacuous.
- c) The union of two intersecting nodal sets is a nodal set.

Proof. Both a) and b) are simple consequences of the definition. Suppose that A_1 and A_2 are intersecting nodal sets and that two distinct points $p, q \in (A_1 \cup A_2) \cap \overline{(M - (A_1 \cup A_2))}$. Since $A_1 \cap \overline{(M - A_1)}$ and $A_2 \cap \overline{(M - A_2)}$ each consist of a single point we must have $\{p\} \cup \{q\} = (A_1 \cup A_2) \cap \overline{(M - A_1)} \cap \overline{(M - A_2)}$. From b), however, it is clear that the intersection of a connected set and two nodal sets is connected, and we have a contradiction.

1.2. In order that the closed set B be a central set it is necessary and sufficient that $\overline{M - B}$ be the union of a finite number of nodal sets, or, equivalently, that B be the intersection of a finite number of nodal sets.

Proof. The necessity is clear. By 1.1 c) any finite union of nodal sets can be reduced to a finite union of non-intersecting nodal sets and the sufficiency follows. From 1.2 we have immediately:

- 1.3. The intersection of a finite number of central sets is a central set.

If each of a system of sets in M has the property that its intersection with a continuum is a continuum or vacuous, then the intersection of all the sets of the system also has this property. Hence,

⁶ See *Mathematische Zeitschrift*, vol. 24 (1926), p. 129.

⁷ The definition of nodal set is due to G. T. Whyburn, *Duke Mathematical Journal*, vol. 4 (1938), p. 2.

1.4. If C is a continuum and A is the intersection of central sets then $A \circ C$ is a continuum or vacuous.

*Definitions.*⁸ A point $p \in M$ is conjugate to $q \in M$ if no point separates p and q in M . If $p \in M$, M_p is defined to be the set of all points conjugate to p . A point $p \in M$ is an *end point* of M if there exists an arbitrarily small neighborhood of p having as its boundary a single point. A set is said to be a simple link provided

a) it consists of an end point or a cut point of M , or

b) it is a *non-degenerate* M_p , where p is a non-cut point.

The latter sets are called *true* simple links.

1.5. Any set M_p is the monotone intersection of central sets.

Proof. The set M_p is clearly the product of all nodal sets containing p in their interior. This product can be written as the product of a countable number of nodal sets, and every partial product is a central set. The result follows.

2. Simple links. From the preceding section we have at once the following:

2.1. THEOREM. A simple link is a continuum. The intersection of a simple link and a continuum is a continuum or vacuous.

2.2. THEOREM. If p_1 and p_2 do not belong to M_q and cannot be joined by a continuum in $M - \{q\}$ then M_q separates M between p_1 and p_2 .

Proof. Write $M_q = \circ B_n$, where $\{B_n\}$ is a monotone sequence of central sets each containing q in its interior. For each $p \in M - M_q$ let N_p be the union of the components about p of $\overline{M - B_n}$. For each p , $N_p \cup M_q$ is closed, for if N_p intersects a component of $\overline{M - B_n}$ it contains this component. If p_1 and p_2 are the points of the hypothesis of the theorem, the sets N_{p_1} and N_{p_2} are disjoint, for p_1 can be joined to any point of N_{p_1} by a continuum in $M - \{q\}$, and similarly for N_{p_2} . Finally, almost all the sets N_p lie within any B_n , since $\overline{M - B_n}$ has a finite number of components. Hence, $M - M_q = (N_{p_1} - M_q) \cup (M - N_{p_1} - M_q)$ is a separation between p_1 and p_2 .

⁸ See K. Kuratowski and G. T. Whyburn, *Fundamenta Mathematicae*, vol. 16 (1930), pp. 305-331. R. L. Moore defines the sets M_p (ref. 4).

2.3. THEOREM. *If p is neither a cut point nor an end point then there exists a point $q \neq p$, q conjugate to p , i. e., M is the union of its simple links.*

Proof. Suppose that p has no conjugate point and that it is not a cut point. Write $\{p\} = M_p = \alpha B_n$, where B_n is a monotone decreasing sequence of central sets. For $\epsilon > 0$ choose m so that $\text{dia } B_m < \epsilon$. Since p is not a cut point and $M_p = \{p\}$ we may, by the previous theorem, join the components of $\overline{M - B_m}$ to each other by continua in $M - \{p\}$. Let A be a continuum $p \in A$, $A \supset \overline{M - B_m}$. For m_1 sufficiently large, $B_{m_1} \cap A = \emptyset$ and the complement of the component about A of $\overline{M - B_{m_1}}$ is a neighborhood of p of diameter $< \epsilon$ with a single boundary point. It follows that p is an end point.

2.4. THEOREM. *Each non-cut point of M belongs to one and only one simple link.*

Proof. We need only consider $p \in M_q \cap M_r$, where M_q and M_r are true simple links. If $\overline{M_q - M_r} \neq \emptyset$ there exists a nodal set A containing M_q but not M_r . Then $\overline{M - A}$ is a nodal set containing M_r and $A \cap \overline{M - A}$ must be p , and p is then a cut point.

2.5. THEOREM. *In order that two points p and q belong to the same simple link it is necessary and sufficient that p and q be conjugate.*

Proof. If p and q belong to the same simple link they are surely conjugate. Suppose p is conjugate to q , $p \neq q$. Let $A = M_p \cap M_q$. Now A is a continuum, and since M_p can be separated in M by no point save possibly p , and similarly for M_q , it follows that A cannot be separated in M by any point. No cut point in A can then be of order 2, and hence⁹ we can choose a non-cut point $r \in A$. Clearly, $p, q \in M_r$ and the theorem is proved.

2.6. THEOREM. *If a point p does not belong to a true simple link it is a regular point in the sense of Menger-Urysohn.*

Proof. Write $\{p\} = M_p = \alpha B_n$, where $\{B_n\}$ is a monotone sequence of central sets, each containing p in its interior. Then $\{B_n - \overline{M - B_n}\}$ is the required sequence of neighborhoods.

Since no cut point belonging to a true simple link can be of order 2 we have

⁹ The order of a point p is the least integer n such that there exists an arbitrarily small neighborhood of p with at most n boundary points. See K. Menger, *Kurventheorie*, p. 96. All but a countable number of cut points of a continuum are of order 2. See G. T. Whyburn, *Transactions of the American Mathematical Society*, vol. 30 (1928), pp. 597-609.

2.7. THEOREM. *There are in M only a countable number of cut points belonging to true simple links.*

2.8. THEOREM. *In order that a non-degenerate closed subset of M be a true simple link it is necessary and sufficient that it be separated in M by no point of M , and that it be saturated relative to this property.*

Proof. The necessity is clear. Suppose that A is saturated relative to the property. If $p, q \in A$, p is conjugate to q and hence by 2.5, $p, q \in M_r$, a true simple link. Since any point not in M_r fails to be conjugate to one of p or q , $A \subset M_r$. Since M_r is not separated in M by any point of M , $M_r = A$.

2.9. THEOREM. *In order that a closed set $A \subset M$ be a continuum it is necessary and sufficient that $A \circ F$ be a continuum for every simple link F and that A contain every point separating a pair of its points in M .*

Proof. The necessity is clear. Let A be a closed set satisfying the two given conditions and suppose $A = A_1 \cup A_2$ is a separation. Let A^* be a continuum irreducible with respect to the property of cutting both A_1 and A_2 . Then A^* must be irreducible about two points $p \in A_1 \circ A^*$, $q \in A_2 \circ A^*$. If p is conjugate to q and F is the simple link about p and q then $A \circ F$ is connected and $A_1 \cup A_2$ cannot be a separation between p and q . If $r \in M$ separates p and q in M then $r \in A^*$ and A^* cannot be irreducible between A_1 and A_2 . We have a contradiction.

PROBLEM. It is of interest to determine the relation between the simple link decomposition and Whyburn's E_0 -set decomposition. Each simple link F is itself a continuum and we may decompose F into simple links $\{F^1\}$ in precisely the same way we decompose M . Proceeding inductively we define F^α for every ordinal α .

- 1) if α has a predecessor $\alpha - 1$ the sets F^α are the simple links of $\dots F^{\alpha-1}$ sets.
- 2) if α is a limit number the sets F^α are all possible intersections $\circ F^\beta$, where this product contains a set from every class $\{F^\beta\}$, for $\beta < \alpha$.

Now a set is both an F^α and an $F^{\alpha+1}$ if and only if it contains no cut point of itself. Any E_0 -set is a subset of some F^α for every α . Hence, any $E_0 = F^\beta$ for some β in the first or second number class. (We can then deduce that the product of an E_0 -set and a continuum is a continuum or vacuous.) The problem is

For a given M is there a number $\beta = \beta(M)$ in the first or second number class such that the class of sets F^β is identical with the class of E_0 -sets?

3. Generalizations of A -sets; Cyclically reducible and extensible properties. In a manner analogous to a characterization of an A -set for (see ref. 8) Peano continua we define

Definition. A subcontinuum of M is a J -set if it is the union of simple links.

We remark that not all J -sets are intersections of central sets.

We have immediately from 2.9 and 2.1 the

3.1. THEOREM. *The intersection of a continuum with a J -set is a continuum or vacuous.*

The following theorems are now obvious.

3.2. THEOREM. *The intersection of any number of J -sets is a J -set, or vacuous.*

3.3. THEOREM. *A J -set contains all the irreducible continua about any two of its points.*

In generalizing the notion of cyclic chain it is most natural to define the cyclic chain $C(a, b)$ as the intersection of all J -sets containing a and b . However, much of the usefulness of the cyclic chain is lost. If $E(a, b)$ is the set of all points separating a and b in M , then every true simple link in $C(a, b)$ has at least one point in common with the closure of $E(a, b) \cup \{a\} \cup \{b\}$; there may be only one common point. Every true simple link in $C(a, b)$ separates a and b in M ; there may be, however, cut points in $C(a, b)$ belonging to no true simple link of $C(a, b)$ which fail to separate a and b in M . A characterization of $C(a, b)$ in terms of $E(a, b)$ is too complicated to be useful.

We define in the usual way (see ref. 8)

Definition. A property is *simple link extensible* if when every simple link of M has the property so has M . A property is *simple link reducible* if when M has the property so also has every simple link of M .

3.4. THEOREM. *Unicoherence is simple link extensible (but not necessarily simple link reducible).*

Proof. If $M = M_1 \cup M_2$ where M_1 and M_2 are continua and if every simple link is unicoherent, then $M_1 \cap M_2$ satisfies the hypothesis of 2.9 and

is hence connected. On the other hand the plane set described in polar coordinates by $\rho = 1$ for $0 \leq \theta \leq 2\pi$ and $\rho = 1 + 1/\theta$ for $\theta \geq 1$ is a unicoherent continuum containing a non-unicoherent simple link.

The relation of the space to its simple links is naturally more complicated in general than is the case for Peano continua. By means of examples it can be shown that

3. 5. a) *A simple link (E_0 -set) is not necessarily a retract of the space.*¹⁰

b) *The fixed point property for homeomorphisms (i. e., the property that under any homeomorphism of the space into itself at least one point remains fixed) is not simple link reducible.*

However, the following will be a direct consequence of 4. 8:

3. 6. THEOREM. *The fixed point property for homeomorphisms is simple link extensible.*

4. Fixed sets under continuous mappings. Throughout this section T will denote a given continuous map of M into M . The following has been shown by the author in a previous paper (see ref. 2).

4. 1. *If $T: M \rightarrow M$ is continuous then $\pi = \cap T^i(M)$, $i = 1, \dots$, is a non-vacuous continuum fixed under T (i. e., $T(\pi) = \pi$).*

4. 2. *The property of a closed set A of being contained in its transform $T(A)$ is inducible.*

Proof. Suppose that $A = \cap A_i$, where $T(A_i) \supset A_i$ and $A_i \supset A_{i+1}$. If $p \in A$ then since $p \in A_i$ for all i and $T(A_i) \supset A_i$ it must be true that $T^{-1}(p) \cap A_i \neq \emptyset$ for all i . Hence, $T^{-1}(p) \cap A \neq \emptyset$ and $p \in T(A)$.

Definitions. A continuum $A \subset M$ has property P if

a) it is a central set, and

b) if B is a component of $\overline{M - A}$, with $B \cap (\overline{M - A}) = \{p\}$, then $T(p) \in B - \{p\}$.

Notice that M itself has P , since any component of $\overline{M - M}$ surely satisfies b).

A continuum in M has property P^* if it is the intersection of sets having property P . The property P^* is clearly inducible.

4. 3. a) *If A_1 and A_2 have P and $A_1 \cap A_2 \neq \emptyset$ then $A_1 \cap A_2$ has P .*

¹⁰ A set $A \subset B$ is a retract of B if there is a continuous map of B onto A which is the identity on A . See K. Borsuk, *Fundamenta Mathematicae*, vol. 17 (1931), pp. 152-170.

- b) If A is irreducible with respect to the property P^* then A is a simple link. If A is a true simple link it may, by a), be written as the intersection of a monotone family of sets having P .

Proof. We first prove a). Suppose that B_1 and B_2 are nodal sets with $\{p_i\} = B_i \cap \overline{M - B_i}$, $T(p_i) \in B_i - \{p_i\}$, ($i = 1, 2$) and $B_1 \cap B_2 \neq \emptyset$. Then by 1.2c), $B_1 \cup B_2$ is nodal and we can suppose that $\{p_1\} = (B_1 \cup B_2) \cap \overline{M - (B_1 \cup B_2)}$. If $p_2 \neq p_1$ then $B_2 \subset B_1$ and $T(p_1) \in B_1 \cup B_2 - \{p_1\}$. If $p_1 = p_2$, clearly we again have $T(p_1) \in B_1 \cup B_2 - \{p_1\}$. Making use of this addition of nodal sets and of 1.4 the truth of a) is clear.

We now prove b). Suppose that $A = \cap A_i$ is a set irreducible with respect to P^* , and that $\{A_i\}$ is a sequence of sets with P . If B is any nodal set, then either B or $\overline{M - B}$ has P . If B is nodal and has P , and $B \cap A \neq \emptyset$ then $B \cap A = \cap (B \cap A_i)$ has P^* . Hence, A is separated in M by no point of M and is therefore a subset of a simple link. But A is the intersection of central sets and must therefore be a simple link.

Definition. For $A \subset M$ we define $A_1 = A \cap T(A)$ and inductively $A_n = A \cap T(A_{n-1})$. Let $D(A) = \cap A_n$, $n = 1, \dots$. If $A_0 \subset A$ clearly $D(A_0) \subset D(A)$.

4.4. For any A having P , $D = D(A)$ is a non-void continuum and $T(D) \supset D$.

Proof. Write $\overline{M - A} = \cup B_i$, $i = 1, \dots, n$, where B_i are pairwise disjoint continua with $B_i \cap A = \{p_i\}$, $T(p_i) \in B_i - \{p_i\}$. For $p \in A$ we define a map $\phi(p)$ as follows

$$\begin{aligned} \text{if } T(p) \in A \text{ then } \phi(p) &= T(p) \\ \text{if } T(p) \in B_i \text{ then } \phi(p) &= p_i. \end{aligned}$$

The continuity of ϕ is readily verified. Now $\phi(A) = A_1$, and it is a simple induction proof to show that $\phi^n(A) = A_n$. By 4.1, $D = \cap \phi^n(A)$ is a continuum fixed under ϕ . We must show that $T(D) \supset D$. If D consists of a single point, in view of the definition of ϕ , this point is surely fixed under T . If D is non-degenerate $T(D) \supset D - \cup \{p_i\}$, $i = 1, \dots, n$, and hence $T(D) \supset D$.

4.5. THEOREM. If $T: M \rightarrow M$ is a continuous transformation then there exists a continuum π , a subset of a simple link of M , such that $T(\pi) \supset \pi$.

Proof. Let $A = \cap A_i$ be irreducible with respect to P^* where $\{A_i\}$ is a monotone decreasing sequence of sets with P . Let $\pi = \cap D(A_i)$. The sequence $\{D(A_i)\}$ is monotone decreasing and $T[D(A_i)] \supset D(A_i)$. The theorem then follows from 4.2 and 4.3 b).

If π is degenerate we have a fixed point. Hence

4.6. THEOREM. *If $T: M \rightarrow M$ is continuous, there is either a fixed point in M or a simple link F such that $F \circ T(F)$ is a non-degenerate continuum.*

4.7. THEOREM. *If $T: M \rightarrow M$ is continuous there exists a compact subset A of a simple link of M such that $T(A) = A$.*

Proof. Let π be the set of 4.5. Define $A_1 = \pi \circ T^{-1}(\pi)$, and then inductively for each n , $A_n = \pi \circ T^{-1}(A_{n-1})$. If $A_{n-1} \subset A_{n-2}$ then $A_n = \pi \circ T^{-1}(A_{n-1}) \subset \pi \circ T^{-1}(A_{n-2}) = A_{n-1}$ and hence $\{A_n\}$ is a monotone sequence of sets. For any $p \in \pi$, $T^{-1}(p) \circ \pi \neq 0$ and hence no $A_n = 0$.

Let $A = \bigcap A_n$. Then A is a compact subset of a simple link. If now $p \in A$, $T(p) \in A_n$ for all n and hence $T(p) \in A$. Therefore $T(A) \subset A$. Furthermore, if $p \in A$, $T^{-1}(p) \circ \pi$ is contained in every A_n , and hence in A , and $p \in T(A)$. Therefore $T(A) \supset A$ and the theorem is proved.

As a corollary to 4.5 and 2.4 we have

4.8. THEOREM. *If $T: M \rightarrow M$ is continuous, and if T carries each simple link into a subset of a simple link—if, for example, the inverse of no point separates a simple link in M —then there exists a simple link F such that $T(F) \subset F$.*

For Peano continua we have

4.9 THEOREM. *If $T: M \rightarrow M$ is a continuous transformation of a Peano continuum into itself then*

- a) *there exists a subcontinuum π of a cyclic element such that $T(\pi) \supset \pi$.*
- b) *either there is a fixed point or some cyclic element intersects its image in a non-degenerate continuum.*
- c) *there is a compact subset A of a cyclic element which is fixed under T .*
- d) *If T carries each cyclic element into a subset of a cyclic element—if, for example, the inverse of no point separates a cyclic element—then T carries some cyclic element into a subset of itself.*
- e) (Ayres (see ref. 1)) *if T is a homeomorphism T carries some cyclic element into a subset of itself.*
- f) (Scherrer (see ref. 6)) *if M is a dendrite then there is a fixed point under T .*

INEQUALITIES OF HIGHER DEGREE IN ONE UNKNOWN.*

By BRUCE ELWYN MESERVE.

1. Introduction. The determination of the content of linear systems of inequalities has been discussed in detail (see [1] for a bibliography), whereas the corresponding problem for non-linear systems has received little consideration, the object of such a work as [3] being to prove inequalities rather than to solve them. The main purpose of the present paper is to discuss the content of the system which asserts that each member of a finite set of polynomials in a single indeterminate is positive.

A system S composed of a finite number of inequalities of the form

$$(1.1) \quad 0 < f;$$

where f is a real-valued function of real variables x_1, x_2, \dots, x_m , has as *content* all sets of real x 's (or points) for which (1.1) is satisfied. If f is continuous, the content of (1.1) is an open set composed of regions whose boundaries are contained in the set of real zeros of f .

Here we are concerned only with the case where f is a polynomial in a single indeterminate. The content then consists of a finite number (possibly zero) of non-overlapping (possibly abutting) segments $\alpha < x < \beta$.

The present paper determines (12) the number of segments $K(a, b)$ in the content of a finite system

$$0 < g_i(x) \quad (i = 1, 2, \dots, s)$$

in $a < x \leq b$, where a, b are real numbers or $\pm \infty$. It also gives (8) bounds for $K(a, b)$, in particular, a bound in terms of $k_1(a, b), \dots, k_s(a, b)$ which denote the number of segments in the contents of the individual inequalities. There is no restriction on the polynomials, the case of common zeros being handled (9) by rational factorization.

The most noteworthy previous paper which deals with the subject and which has come to our attention is by C. F. Gummer [2]. He gives a method for determining what we denote by $K(-\infty, \infty)$ under certain conditions

* Received August 13, 1946; Presented to the American Mathematical Society, August 20, 1946.

[2; 268, 278], the most restrictive of which from the standpoint of inequalities is that the zeros be simple. His process being primarily designed to determine the relative order of the real zeros of the polynomials, his treatment of inequalities is somewhat incidental but clearly brings out the fundamental importance of the Cauchy-Sylvester theorem for the theory. The present paper employs a generalized form of that theorem to give a different process which is both simpler and unrestricted.

An elegant treatment of mixed systems has been given by A. Markoff in *Recueil Mathématique*, vol. 7 (1940), pp. 3-6.

2. Segments and their associated polynomials. Let (α, β) denote a segment (open) and $[\alpha, \beta]$ an interval (closed). For our purposes the general linear inequality can be regarded as having one of the forms

$$(2.1) \quad 0 < x - \alpha \quad \text{or} \quad 0 < -x + \beta$$

with the segments (α, ∞) , $(-\infty, \beta)$ for respective contents. Similarly, the general quadratic inequality has one of the forms

$$(2.2) \quad 0 < -x^2 + 2bx - c \quad \text{or} \quad 0 < x^2 + 2bx + c$$

with $(\alpha, \beta) + (\beta, \alpha)$ and the complement of $[\alpha, \beta] + [\beta, \alpha]$ for respective contents, where α, β are the zeros of the corresponding right member and (α, β) , $[\alpha, \beta]$ are to be interpreted as vacuous except when α, β are real numbers satisfying $\alpha < \beta$, $\alpha \leq \beta$ respectively. From the standpoint of content (2.1) can be regarded as quadratics of the first type in (2.2) if their right members are regarded as having the real zeros α, ∞ and $\beta, -\infty$ respectively.

Conversely, with the exception of the whole line of points, every segment (α, β) , where α, β are distinct real numbers or $\pm \infty$ is the content of a unique inequality of one of the three types: $0 < x - \alpha$; $0 < -x + \beta$; $0 < -x^2 + 2bx - c$, with $0 < b^2 - ac$. The segment will be said to *belong* to the corresponding polynomial; the polynomial to *represent* the segment. In general a polynomial of degree n is positive for interior points and zero for boundary points of each of its segments.

3. Content of a polynomial inequality. Let $f(x)$ be a polynomial with real coefficients. The real zeros of f divide the line into a finite number of segments on which f has constant sign. Let $\pi_1, \pi_2, \dots, \pi_k$ be the poly-

nomials representing the segments on which f is positive. Since the boundaries of the π 's are zeros of f , we have

$$(3.1) \quad f(x) = \sigma \pi_1 \pi_2 \cdots \pi_k.$$

If $0 < f(x)$, one of the π 's is positive, the $k-1$ others are negative and σ has the same sign as $(-1)^{k-1}$. If $0 < -f(x)$, all of the π 's are negative and again σ has the sign of $(-1)^{k-1}$. The polynomial σ vanishes at all zeros of f except simple zeros and zero minima of multiplicity exactly two.

The π 's can be determined theoretically by finding the real zeros of f . From a knowledge of the real zeros, their multiplicities and the sign of the initial of f , the graphical representation of the segments is rapidly made. A zero x of even multiplicity is a zero minimum if and only if

$$(3.2) \quad 0 < (-1)^{N(x, \infty)} \operatorname{sgn} a_0,$$

where $N(x, \infty)$ is the number of real zeros greater than x , each zero being counted a number of times equal to its multiplicity, and a_0 is the initial of f .

4. The number of segments. Our main problem is to determine k , which is the number of segments on which $f(x)$ is positive. A first step in this direction is given by

THEOREM 4.1. *The number k of segments on which $f(x)$ is positive is given by*

$$k = \frac{1}{2}\{N_0 + e(-\infty) + e(\infty)\} + \mu,$$

where

$$e(x) = 0, \text{ if } f(x) \leq 0, \quad e(x) = 1, \text{ if } 0 < f(x),$$

N_0 is the number of real zeros of $f(x)$ of odd multiplicity, and μ is the number of zero minima of $f(x)$.

The number of segments which belong to f is given by the same formula.

Bounds for the number of segments are given by

$$\frac{1}{2}\{N_0 + e(-\infty) + e(\infty)\} \leq k \leq \frac{1}{2}\{N_0 + e(-\infty) + e(\infty)\} + N_e,$$

where N_e is the number of real zeros whose multiplicity is even, and by

$$(4.1) \quad 0 \leq k \leq \left[\frac{m}{2} \right] + 1,$$

where $m = \deg f$ and the brackets denote "the greatest integer in."

Similar formulas hold for counting the number of segments $k(a, b)$ on which f is positive and which are contained in the point set

$$(4.2) \quad a < x \leq b.$$

For example,

$$(4.3) \quad k(a, b) = \frac{1}{2}\{N_0(a, b-0) + e(a+0) + e(b-0)\} + \mu(a, b-0),$$

where $e(a+0)$ denotes the limit from the right and $e(b-0)$ the limit from the left. Either or both of the extreme segments counted in this formula may fail to *belong to* f because f may not vanish everywhere on the boundary.

Unlike $N_0(a, b)$, the function $k(a, b)$ is not an additive function of (4.2) because an end segment does not necessarily belong to the polynomial f .

The numbers N_0 , N_e can be determined by the Sturm sequences [4]. On the other hand, the Sturm sequences for

$$f(x) = (x-1)(x-2)^2,$$

which has a zero minimum, and for

$$g(x) = (x-1)^2(x-2),$$

which does not have a zero minimum, have identical signs at $-\infty$ and at $+\infty$. Thus the zero minima cannot be counted by means of the signs of the Sturm polynomials for f alone. In the general case something like the Cauchy-Sylvester theorem (6) seems necessary for the separation of the zero maxima from the zero minima.

5. Number of zero extrema. By rational operations, including differentiation, the polynomial f can be written in the form

$$(5.1) \quad f = a_0 f_1 f_2^2 \cdots f_r^r,$$

where a_0 is the initial of f , each f_i has simple roots and initial 1, and each pair of f_i 's is relatively prime. N_0 and N_e may be determined from the Sturm sequences of f_1, f_2, \dots, f_r . N_0 is the total number of real zeros of f_1, f_3, \dots , and N_e is the total number of real zeros of f_2, f_4, \dots .

The zero minima of multiplicity $2h$ are the roots of the *mixed system*

$$(5.2) \quad 0 = g_1, \quad 0 < g_2,$$

where

$$g_1 = f_2^h, \quad g_2 = f_1 f_3 \cdots,$$

the zeros of g_2 being the zeros of f with odd multiplicity. The next section will show how to count the roots of (5.2).

6. The fundamental lemmas. The Cauchy-Sylvester theorem employed by Gummer [2] can be generalized to cover the case of multiple zeros in exactly the same way as Sturm's theorem [4]. Let g_1, g_2 be two relatively prime polynomials with real coefficients (a similar process and result are readily seen to apply when g_1, g_2 have no common real zero on $a < x \leq b$).

Form the division sequence

$$(6.1) \quad H_1, H_2, \dots, H_q,$$

in which

$$(6.2) \quad \begin{aligned} H_i &= H_{i+1}Q_{i+1} - H_{i+2} & (i = 1, 2, \dots, q-1), \\ H_1 &= g_1, \quad H_2 = g'_1g_2, \quad H_{q+1} = 0. \end{aligned}$$

Note that in general H_2 is of higher degree than H_1 , so that H_3 is usually $-g_1$.

First suppose that H_q is a non-zero constant. Relations (6.2) can be used to prove

(i) The number of variations of sign in (6.1) can change only at a zero of H_1 .

This result can be applied to the subsequence obtained by omitting the first term of (6.1) because that subsequence is generated by applying the division process to H_2, H_3 which are relatively prime in the case being considered. Hence the number of variations of sign in the subsequence can change only at a zero of H_2 . Since no zero of H_1 is a zero of H_2 , we have

(ii) The number of variations of sign in

$$H_2, H_3, \dots, H_q$$

does not change at a zero of g_1 .

As x increases through a zero of g_1 , the polynomial g_1^2 decreases to zero and then increases. Hence $g_1g'_1$ is negative just before a zero of g_1 and positive just after. Therefore H_1H_2 has opposite sign to g_2 just before the zero and the same sign as g_2 just after. Hence

(iii) Sequence (6.1) loses one variation of sign at a zero of g_1 if g_2 is positive and gains one variation of sign just after a zero of g_1 if g_2 is negative.

Let $F(x)$ be the number of variations of sign in (6.1) for the real number x . On $a < x \leq b$ let P, N be the number of zeros of g_1 at which g_2 is respectively positive and negative. Then

$$P - N = F(a) - F(b + 0).$$

Except when $g_2(b) < 0 = g_1(b)$, $F(b + 0) = F(b)$.

Now suppose that H_q is not necessarily constant. Divide (6.1) by H_q to give

$$(6.3) \quad K_1, K_2, \dots, 1,$$

a sequence which can be generated from K_1, K_2 in exactly the same way as (6.1) was generated from H_1, H_2 . The zeros of K_1 are simple and are the distinct zeros of g_1 . From $H_1 H_2 = H_q^2 K_1 K_2$ we see that the behavior of $K_1 K_2$ near a zero of g_1 is the same as that of $H_1 H_2$. Denote the number of variations of sign in the sequence (6.3) by $F_{11}(x)$.

If $H_q(x)$ is not constant, form the division sequence for $H_q, H'_q g_2$, remove the highest common factor from its terms and denote the number of variations by $F_{12}(x)$. The zeros of the first term are simple and are the zeros of g_1 with multiplicity at least two.

If the highest common factor of $H_q, H'_q g_2$ is not constant, form a third sequence with $F_{13}(x)$ variations; and so on.

Put

$$G_{1j} = F_{1j}(a) - F_{1,j+1}(a) - F_{1j}(b + 0) + F_{1,j+1}(b + 0).$$

By taking $g_2 = 1$, we find Sturm sequences for g_1 . Denote by S_{1j} the functions corresponding to G_{1j} , the argument $b + 0$ being replaced by b .

THEOREM 6.1. *If g_1, g_2 are relatively prime polynomials with real coefficients, the numbers P_{1j}, N_{1j} of zeros of g_1 which lie on $a < x \leq b$, which have multiplicity j and which make g_2 respectively positive and negative satisfy*

$$P_{1j} + N_{1j} = S_{1j}, \quad P_{1j} - N_{1j} = G_{1j}.$$

Theorem 6.1 and the corresponding generalized Sturm theorem which it implies will be called the fundamental lemmas. From them, Theorem 4.1, and (4.3) we see that the number of non-overlapping segments on which $0 < f(x)$ can be determined by rational operations.

7. The resolvent. Consider now a finite system S

$$(7.1) \quad 0 < g_i \quad (i = 1, 2, \dots, s),$$

where the g 's are polynomials with real coefficients and a single indeterminate x . Since the intersection of two open sets is an open set, the content of S is a finite number of non-overlapping segments. The system S has the same content as a system containing a single inequality $0 < G$, which will be called a *resolvent* of S .

Various graphical methods for finding the segments of G from those for the g 's can be devised. One such is to represent the number x by a point on a semi-circular arc at θ radians from the mid-point of the arc, where $x = \tan \theta$. Erase the segments belonging to $-g_1, -g_2, \dots, -g_s$ and the boundary points of g_1, g_2, \dots, g_s . The remaining segments are the content of S .

The number of segments which belong to $a < x \leq b$ and on which G is positive will be denoted by $K(a, b)$. The number of segments on which each g_i is positive will be denoted by $k_i(a, b)$. Usually the (a, b) will be omitted from this notation, but it will nevertheless be understood.

A segment belongs to S if and only if it belongs to a resolvent of S , that is, if and only if every polynomial of S is positive on the segment and at least one polynomial of S vanishes at each point of the boundary of the segment. The phrase a "segment of S ," on the other hand, means simply a segment on which G is positive.

8. Bounds on the number of segments. The boundary of the resolvent of S is a subset of the sum of the boundaries of the segments of the polynomials in S . This is merely a restatement of the fact that the real zeros of G are found among those of the product

$$\prod_{i=1}^s g_i.$$

Hence by (4.1) the number $K(-\infty, \infty)$ of segments on which G is positive satisfies

$$(8.1) \quad K(-\infty, \infty) \leq \left[\frac{1}{2} \sum_{i=1}^s m_i \right] + 1,$$

where $m_i = \deg g_i$.

It is convenient for the present count to regard an infinite segment as having two end points so that the number of end points is always double the number of segments. A left end point of g_1 will be an end point of the system $0 < g_1, 0 < g_2$ if and only if it is either a left end point or an interior point of a segment of g_2 . Denote by i_1 the number of end points of g_1 which are interior points of segments of g_2 , by i_2 the number of end points of g_2 which are interior points of segments of g_1 , by λ the number of points which

are left end points for both g_1 and g_2 , and by ρ the number of points which are right end points for both g_1 and g_2 . The total number of end points for $0 < g_1, 0 < g_2$ is then $i_1 + i_2 + \lambda + \rho$ and the number K of segments is given by

$$(8.2) \quad 2K = i_1 + i_2 + \lambda + \rho.$$

If two end points coincide, by shifting one of them slightly we replace the system by another with the same K, k_1, k_2 but with the corresponding pair of end points distinct. Hence in determining formulas for K we may assume that no end points coincide. Similarly, we may arrange that all end points are finite. For the modified system we have

$$2K = i_1 + i_2,$$

$$2k_1 = i_1 + j_1, \quad 2k_2 = i_2 + j_2,$$

where j_1 is the number of end points of g_1 which are exterior to the segments of g_2 and j_2 is analogously defined. Accordingly,

$$K = k_1 + k_2 - \frac{1}{2}(j_1 + j_2).$$

Now the end point farthest to the left is exterior to the segments of g_1 or to the segments of g_2 . Similarly for the end point farthest to the right. Hence the least value of $j_1 + j_2$ is 2 and

$$K \leq k_1 + k_2 - 1.$$

The bound can be attained for all positive integral values of k_1, k_2 . This is obvious for $k_1 = k_2 = 1$. Assume the result for k_1, k_2 and add a segment which covers one of the extreme end points. Both K and k_1 (say) are increased by 1 so that the relation

$$K = k_1 + k_2 - 1$$

continues to hold. By $s - 1$ applications of the above we deduce

THEOREM 8.1. *A finite system S (7.1) has K segments, where*

$$(8.3) \quad K \leq \sum_{i=1}^s k_i - s + 1$$

and k_i denotes the number of segments of g_i . This bound can be attained.

9. Reduction of degree. Two systems having the same content are equivalent.

The *degree* of a system of inequalities is the sum of the degrees of the polynomials appearing in the system. Equivalent systems do not necessarily have the same degree.

Let a highest common factor of f, g be h and set

$$f = f^*h, \quad g = g^*h.$$

Let $S^{(1)}, S^{(2)}$ be the systems found by replacing in S the two inequalities $0 < f, 0 < g$ by the three inequalities

$$0 < f^*, \quad 0 < g^*, \quad 0 < h$$

and by the three inequalities

$$0 < -f^*, \quad 0 < -g^*, \quad 0 < -h$$

respectively. We may then write

$$S = S^{(1)}S^{(2)}$$

in the sense that every solution of $S^{(1)}$ or $S^{(2)}$ is a solution of S and every solution of S is a solution of $S^{(1)}$ or a solution of $S^{(2)}$. Moreover, $S^{(1)}$ has no solution in common with $S^{(2)}$, that is, no segment of $S^{(1)}$ intersects a segment of $S^{(2)}$. Consequently $K = K_1 + K_2$, where K_i denotes the number of segments of $S^{(i)}$.

Put

$$\deg f^* = m^*, \quad \deg g^* = n^*, \quad \deg h = p.$$

Then

$$\deg S = m^* + n^* + 2p + q,$$

$$\deg S^{(1)} = \deg S^{(2)} = m^* + n^* + p + q.$$

Hence, if p is positive,

$$\deg S^{(1)} < \deg S, \quad \deg S^{(2)} < \deg S.$$

Since the degree of $S^{(i)}$ is finite and non-negative, we have by repetition of the above process

THEOREM 9.1. *A finite system S may be reduced by rational operations to the product of a finite number of systems such that each pair of polynomials in each factor is relatively prime and such that the number of segments of S equals the sum of the numbers of segments of the factors.*

10. **The general system in one indeterminate.** In connection with the system S

$$0 < g_1, \quad 0 < g_2, \quad \dots, \quad 0 < g_s$$

we shall have occasion to consider the *mixed systems* S_i got from S by replacing one inequality $0 < g_i$ by the corresponding equation $0 = g_i$. A number satisfying S_i will be called a *root* of S_i and endowed with the multiplicity which it has as a zero of g_i . In the sections immediately following we shall be interested in the number of roots on a point set $a < x \leq b$. In general, however, the (a, b) will be omitted from the notation just as $K(a, b)$ has already been abbreviated to K .

Denote the number of real roots of S_i by P_i and the number of real roots of multiplicity j by P_{ij} so that

$$P_i = \sum_{j=1}^{r_i} P_{ij}.$$

Denote by $\mu_{i,2h}$, $M_{i,2h}$ the numbers of zero minima and zero maxima of S_i with multiplicity $2h$. The total numbers of zero extrema are then given by

$$\mu_i = \sum_{h=1}^{\kappa_i} \mu_{i,2h}, \quad M_i = \sum_{h=1}^{\kappa_i} M_{i,2h},$$

where $2\kappa_i$ is the last even index on a non-vanishing function of the sequence F_{ij} defined in 6.

Define also the number d_i as the number of roots of S_i got by counting each root of odd multiplicity once, counting each zero minimum twice and omitting each zero maximum.

11. **Mixed systems.** Consider now the two mixed systems

$$0 = g_1, \quad 0 < g_2, \quad P_1;$$

$$0 = g_1, \quad 0 < -g_2, \quad N_1;$$

the capital letter designating in each case the number of roots of a given multiplicity whose index j is omitted for convenience. If we set

$$(11.1) \quad P_1 + N_1 = S_1, \quad P_1 - N_1 = G_1,$$

the calculation of P_1 , N_1 is equivalent to that of S_1 , G_1 . The number S_1 is really S_{1j} and is not to be confused with the system S_1 .

More generally, consider the systems

$$\begin{array}{llll}
0 = g_1, & 0 < g_2, & 0 < g_3, & P_1; \\
0 = g_1, & 0 < g_2, & 0 < -g_3, & Q_1; \\
0 = g_1, & 0 < -g_2, & 0 < g_3, & Q_2; \\
0 = g_1, & 0 < -g_2, & 0 < -g_3, & Q_3.
\end{array}$$

Then we have also the systems

$$\begin{array}{lll}
0 = g_1, & 0 < g_2, & P_1 + Q_1; \\
0 = g_1, & 0 < -g_2, & Q_2 + Q_3; \\
0 = g_1, & 0 < g_2 g_3, & P_1 + Q_3; \\
0 = g_1, & 0 < g_3, & P_1 + Q_2.
\end{array}$$

If we assume that the number of roots can be determined for every fixed system with just one inequality, the numbers $P_1 + Q_1$, $Q_2 + Q_3$, $P_1 + Q_3$, $P_1 + Q_2$ can be calculated and from them P_1 , Q_1 , Q_2 , Q_3 .

The same remarks apply to the systems got by adjoining $s - 2$ inequalities to each system above. The general system

$$0 = g_1, \quad 0 < g_2, \quad \dots, \quad 0 < g_s, \quad P_1$$

for $2 < s$ is one of 2^{s-1} mixed systems, all of which contain $0 = g_1$ and which can be grouped in sets of four. The number of their roots can be computed in terms of the numbers for mixed systems with $s - 2$ inequalities. Hence by induction the calculation of P_1 for the general system is referred back to the numbers S_1 and G_1 appearing in (11.1).

12. The number of segments of a finite system. The numbers in (11.1) and consequently all the numbers in 11 can be determined by the fundamental lemmas provided only that every pair of g 's is relatively prime. The factorization of 9 guarantees this for the factors $S^{(1)}, S^{(2)}, \dots$ of S_1 and the numbers for S_1 can be got by adding those for $S^{(1)}, S^{(2)}, \dots$.

The $\mu_{1,2h}$ zero minima of S_1 with multiplicity $2h$ are the roots of multiplicity $2h$ of the mixed system

$$0 = g_1, 0 < g_1^{(2h)}, 0 < g_2, \dots, 0 < g_s.$$

The $M_{1,2h}$ zero maxima of S_1 with multiplicity $2h$ are likewise the roots of multiplicity $2h$ of

$$0 = g_1, 0 < -g_1^{(2h)}, 0 < g_2, \dots, 0 < g_s.$$

The fundamental lemmas, not directly applicable to these because g_1 and $g_1^{(2h)}$ are not relatively prime, become applicable by the process of 5.

If g_1, g_2 are two relatively prime polynomials, no segment of g_1 has an end point in common with a segment of g_2 . The numbers i_1, i_2 in (8.2) are respectively the d_1, d_2 of 10. Nothing is contributed to λ, ρ by coincident zeros in this case, but either λ or ρ may be unity because of end points induced by the end points of the fundamental interval (a, b) . Using the notation of 4 we find the formula

$$K = \frac{1}{2}\{d_1(a, b-0) + d_2(a, b-0) + e_1(a+0)e_2(a+0) + e_1(b-0)e_2(b-0)\}.$$

This formula is easily generalized to

$$K(a, b) = \frac{1}{2}\left\{\sum_{i=1}^s d_i(a, b-0) + \prod_{i=1}^s e_i(a+0) + \prod_{i=1}^s e_i(b-0)\right\}.$$

As in the case $s=1$, the number $K(a, b)$ can be found by rational operations.

13. A sufficient condition for consistency in terms of the resultant.

If the roots of f are ξ_i , those of g are η_j , $\deg f = m$ and $\deg g = n$, the resultant for the order f, g may be defined as

$$R = (-1)^{\frac{1}{2}m(m-1)} a_0^n \prod g(\xi) = (-1)^{\frac{1}{2}m(m-1) + mn} b_0^m \prod f(\eta).$$

If all roots of f are real, then

$$0 < (-1)^{m+1} \prod g(\xi)$$

is a sufficient condition for one of the numbers $g(\xi)$ to be positive. Hence when all the roots of f are real,

$$(13.1) \quad 0 < (-1)^{\frac{1}{2}m(m+1)+1} a_0^n R,$$

where R is expressed in terms of the coefficients of f, g , is a sufficient condition for the consistency of the system

$$(13.2) \quad 0 < f, \quad 0 < g,$$

and when all the roots of g are real

$$(13.3) \quad 0 < (-1)^{\frac{1}{2}m(m-1) + mn + n+1} b_0^m R$$

is a sufficient condition. From subsequent developments (Theorem 14.1) it will be apparent that the condition is by no means necessary.

14. The case of two quadratics. Let us examine the condition of consistency for a system of two inequalities, each of which defines a finite segment, that is, in (13.2) suppose

$$f = -x^2 + 2bx - c, \quad g = -x^2 + 2b'x - c'.$$

Denote the respective segments by (α, β) , (γ, δ) . They have a point in common if and only if

$$(14.1) \quad 0 < (\alpha - \delta)(\gamma - \beta).$$

Multiplication gives

$$0 < \alpha\gamma + \beta\delta - \alpha\beta - \gamma\delta.$$

Setting

$$\alpha = b - d, \quad \beta = b + d, \quad \gamma = b' - d', \quad \delta = b' + d',$$

where

$$0 < d = (b^2 - c)^{\frac{1}{2}}, \quad 0 < d' = (b'^2 - c')^{\frac{1}{2}},$$

we find

$$\alpha\beta = c, \quad \gamma\delta = c', \quad \alpha\gamma + \beta\delta = 2(bb' + dd').$$

Hence condition (14.1) can be written

$$(14.2) \quad 0 < 2(bb' + dd') - c - c'.$$

Let us next obtain an equivalent condition by a method which is applicable when all the zeros are real. We have in the present case for the polynomial whose zeros are the values assumed by g at the zeros of f

$$E(0, g) = -g^2 + 2\{c - c' - 2b(b - b')\}g + R,$$

where R is the resultant

$$R = -(c - c')^2 + 4b(b - b')(c - c') - 4c(b - b')^2.$$

The condition that

$$-x^2 + 2Bx - C \quad (0 \leq B^2 - C)$$

have at least one positive root is $0 < B$ or $0 < -C$. The condition wanted is that at least one of $E(f, 0)$, $E(0, g)$ have at least one positive root. Hence the condition is that one of the following be satisfied:

$$0 < c - c' - 2b(b - b'), \quad 0 < c' - c - 2b'(b' - b), \quad 0 < R.$$

Thus we have

THEOREM 14. 1. *The system of inequalities*

$$0 < -x^2 + 2bx - c = f, \quad 0 < -x^2 + 2b'x - c' = g$$

whose segments are non-vacuous and have no boundary point in common is consistent if and only if

$$T : 0 < 2(bb' + dd') - c - c',$$

where $0 < d = (b^2 - c)^{\frac{1}{2}}$, $0 < d' = (b'^2 - c')^{\frac{1}{2}}$. This system T found by irrational operations may be factored into the product of three systems found by rational operations $T = UVW$, where

$$U : 0 < c - c' - 2b(b - b'), \quad V : 0 < c' - c - 2b'(b' - b),$$

$$W : 0 < R(f, g).$$

DUKE UNIVERSITY.

REFERENCES.

-
- [1] Lloyd L. Dines and N. H. McCoy, "On linear inequalities," *Transactions of the Royal Society of Canada*, (3) vol. 27 (1933), section III, pp. 37-70.
 - [2] C. F. Gummer, "The relative distribution of the real roots of a system of polynomials," *Transactions of the American Mathematical Society*, vol. 23 (1922), pp. 265-282.
 - [3] G. H. Hardy, J. E. Littlewood and G. Pólya, *Inequalities*, Cambridge, 1934.
 - [4] Joseph Miller Thomas, "Sturm's theorem for multiple roots," *National Mathematics Magazine*, vol. 15 (1941), pp. 391-394.

ON RIESZ SUMMABILITY AND SUMMABILITY BY DIRICHLET'S SERIES.*¹

By C. T. RAJAGOPAL.

We know that, if $k > 0$, summability of any function by the k -th Riesz mean necessarily implies its summability by the Laplace transformation. The converse which is not universally true has been discussed by Szász for the case $k = 1$ [5, Theorem 1].² Theorem 1 of this paper is a generalization of Szász's problem for $k \geq 1$. Theorem 2 is an extension of a Tauberian oscillation theorem of V. Ramaswami [3], involving Riesz means and the Laplace transformation in place of his Cesàro means and Abel mean. Theorems t , T are in the nature of corollaries to Theorems 1, 2.

1. THEOREM 1. Suppose that

$$(1) \quad F(t) = t \int_0^\infty A(u) e^{-ut} du \text{ converges for } t > 0,$$

$$(2) \quad F(t) \rightarrow s \text{ as } t \rightarrow +0;$$

$A_r(x)$ defined by

$$A_r(x) = r \int_0^x (x-u)^{r-1} A(u) du, \quad r > 0,$$

$$A_0(x) = A(x),$$

satisfies the condition

$$(3) \quad B_k(x) \equiv (k+1) \{x A_k(x) - A_{k+1}(x)\} \geq -Kx^{k+1} \quad (x > 0),$$

$k \geq 0$, $K > 0$ being constants. Then

$$(4) \quad \sigma_{k+1}(x) \equiv \frac{A_{k+1}(x)}{x^{k+1}} \rightarrow s \text{ as } x \rightarrow \infty,$$

but $\sigma_k(x)$ does not in general tend to a limit.

* Received August 2, 1946.

¹ My thanks are due to Dr. V. Ganapathy Iyer and the referee who have suggested the revision of certain details.

² Numbers in heavy type refer to the list of references at the end of this paper.

1.1. The proof of the theorem requires the following lemmas of which the first and the second are well known and the third is proved by Szász in the course of his demonstration of the case $k = 0$ of the theorem.

LEMMA 1. $F(t)$ defined by (1) satisfies the identity:

$$F(t) = \frac{t^{k+1}}{\Gamma(k+1)} \int_0^\infty A_k(u) e^{-ut} du, \quad k > 0.$$

LEMMA 2. If $s(x) \geq 0$ and

$$t \int_0^\infty s(u) e^{-ut} du \rightarrow K \text{ as } t \rightarrow +0,$$

then

$$\int_0^x s(u) du \sim Kx \text{ as } x \rightarrow \infty.$$

This lemma is a special case of a theorem of Karamata. [6, Satz 2].

LEMMA 3. If

$$(5) \quad \int_0^x (v(u)/u) du = o(x) \text{ as } x \rightarrow \infty,$$

$$(6) \quad t \int_0^\infty \left\{ \int_0^u (v(x)/x^2) dx \right\} e^{-ut} du \rightarrow s \text{ as } t \rightarrow +0,$$

then

$$(7) \quad \int_0^\infty (v(u)/u^2) du = s.$$

LEMMA 4. Suppositions (1) and (2) imply that

$$F_k(t) = t \int_0^\infty \sigma_k(u) e^{-ut} du, \quad k > 0,$$

converges for $t > 0$ and

$$(8) \quad F_k(t) \rightarrow s \text{ as } t \rightarrow +0.$$

Proof. Lemma 1 gives, when $t > 0$,

$$\begin{aligned} \Gamma(k+1) \int_t^\infty \frac{F(x)}{x^{k+1}} (x-t)^{k-1} dx &= \int_t^\infty (x-t)^{k-1} dx \int_0^\infty A_k(u) e^{-ux} du \\ &= \int_0^\infty A_k(u) du \int_t^\infty (x-t)^{k-1} e^{-ux} dx \\ &= \Gamma(k) \int_0^\infty \frac{A_k(u)}{u^k} e^{-ut} du, \end{aligned}$$

the inversion of the order of integration being justified by absolute convergence. Now

$$F(t) - F_k(t) = kt \int_t^\infty [F(t) - F(x)] \frac{(x-t)^{k-1}}{x^{k+1}} dx = kt \int_t^{t^{\frac{1}{2}}} + kt \int_{t^{\frac{1}{2}}}^\infty$$

$t < 1,$

where, as $t \rightarrow 0$,

$$\begin{aligned} kt \int_{t^{\frac{1}{2}}}^\infty F(x) \frac{(x-t)^{k-1}}{x^{k+1}} dx &= O[kt \int_{t^{\frac{1}{2}}}^\infty \frac{(x-t)^{k-1}}{x^{k+1}} dx] \\ &= O[k \int_{x=t^{\frac{1}{2}}}^\infty (1 - (t/x))^{k-1} d(1 - (t/x))] \\ &= O[1 - (1 - t^{\frac{1}{2}})^k]. \end{aligned}$$

Similarly

$$kt \int_t^{t^{\frac{1}{2}}} \frac{(x-t)^{k-1}}{x^{k+1}} dx = (1 - t^{\frac{1}{2}})^k$$

and therefore

$$\begin{aligned} |F(t) - F_k(t)| &< \max_{t \leq x \leq t^{\frac{1}{2}}} |F(t) - F(x)| (1 - t^{\frac{1}{2}})^k \\ &\quad + [1 - (1 - t^{\frac{1}{2}})^k] [|F(t)| + O(1)], \quad t \rightarrow 0, \end{aligned}$$

which, in conjunction with (2), establishes (8).

1.2. Proof of Theorem 1. Lemma 4 for $F_k(t)$ and $F_{k+1}(t)$ can be used in (3) so as to obtain successively

$$\begin{aligned} t \int_0^\infty \frac{B_k(u)}{u^{k+1}} e^{-ut} du &\rightarrow 0 \text{ as } t \rightarrow 0, \\ t \int_0^\infty \left\{ \frac{B_k(u)}{u^{k+1}} + K \right\} e^{-ut} du &\rightarrow K \text{ as } t \rightarrow 0. \end{aligned}$$

Hence we can take, in Lemma 2,

$$s(u) = \frac{B_k(u)}{u^{k+1}} + K$$

and get

$$\int_0^x \left\{ \frac{B_k(u)}{u^{k+1}} + K \right\} du \sim Kx \text{ as } x \rightarrow \infty,$$

or

$$(9) \quad \int_0^x \frac{B_k(u)}{u^{k+1}} du = o(x) \text{ as } x \rightarrow \infty.$$

Also we have, from the definition of $B_k(x)$ in (3),

$$B_k(x) = x^{k+2} \frac{d}{dx} \frac{A_{k+1}(x)}{x^{k+1}}$$

which gives rise to

$$(10) \quad \sigma_{k+1}(x) = \int_0^x \frac{B_k(u)}{u^{k+1}} du.$$

When we write $(k+1)$ in place of k in (8) and use (10), we have a relation which is (6) with $v(u) = B_k(u)/u^k$. The same choice of $v(u)$ makes (5) identical with (9). Lemma 3 therefore enables us to conclude that

$$\int_0^\infty \frac{B_k(u)}{u^{k+2}} du = s, \text{ i. e., } \sigma_{k+1}(x) \rightarrow s \text{ as } x \rightarrow \infty.$$

A necessary condition for the further conclusion $\sigma_k(x) \rightarrow s$ is $B_{k+1}(x)/x^{k+1} \rightarrow 0$ which is more than (3). Hence the further conclusion is not in general warranted.

1.3. In Theorem 1, we can take $A(x) = \sum_{\lambda_n \leq x} a_n$, where $\{\lambda_n\}$ is a positive, increasing, divergent sequence, and obtain

COROLLARY 1. *Suppose that*

$$F(t) = \sum_{\nu=1}^{\infty} a_\nu e^{-\lambda_\nu t}, \quad 0 < \lambda_1 < \lambda_2 < \dots, \lambda_n \rightarrow \infty;$$

$$A_r(x) = \sum_{\lambda_n \leq x} (x - \lambda_n)^r a_n,$$

$$\frac{B_r(x)}{r+1} = \sum_{\lambda_n \leq x} (x - \lambda_n)^r a_n \lambda_n, \quad r \geq 0,$$

Then the conditions:

$$(i) \quad F(t) \text{ converges for } t > 0, \lim_{t \rightarrow +0} F(t) = s,$$

$$(ii) \quad B_k(x) \geq -k\lambda_n^{k+1}, \quad \lambda_n \leq x < \lambda_{n+1}, \quad k \geq 0,$$

together imply that $\sum_{\nu=1}^{\infty} a_\nu$ is summable- $R(\lambda_n, k+1)$, i. e., summable by the Riesz mean of type λ and order $(k+1)$; but not in general summable- $R(\lambda_n, k)$.

This result leads at once to the following extension of a classical theorem of Schnee [2, Theorem 33].

COROLLARY 2. *Necessary and sufficient conditions for the $R(\lambda_n, k)$ -summability ($k \geq 0$) of $\sum_{\nu=1}^{\infty} a_\nu$ are (in the notation of Corollary 1):*

(i) $F(t)$ converges for $t > 0$, $\lim_{t \rightarrow +0} F(t) = s$,

(ii') $B_k(x) = o(\lambda_n^{k+1})$, $\lambda_n \leq x < \lambda_{n+1}$ ($n \rightarrow \infty$).

That the conditions are necessary for the $R(\lambda_n, k)$ -summability of $\sum_{\nu=1}^{\infty} a_{\nu}$ follows from Lemma 1 and the definition of $B_k(x)$.

To prove the sufficiency of the conditions we observe, from Corollary 1, that (i) and (ii') establish the $R(\lambda_n, k+1)$ -summability of $\sum_{\nu=1}^{\infty} a_{\nu}$. The $R(\lambda_n, k)$ -summability of the series follows as a consequence of (ii').

2. When $\lim_{t \rightarrow +0} F(t)$ does not exist, we can (under certain conditions) determine an upper or a lower bound for $\overline{\lim}_{t \rightarrow \infty} \sigma_k(x)$, for large enough k , in terms of $\overline{\lim}_{t \rightarrow +0} F(t)$ or $\lim_{t \rightarrow +0} F(t)$. We have, in this direction,

LEMMA 5. Let us write

$$\overline{\lim}_{t \rightarrow +0} F(t) = \bar{s}, \quad \overline{\lim}_{x \rightarrow \infty} \sigma_k(x) = \bar{\sigma}_k.$$

Let us suppose that $r \geq 0$, $k \geq 1$ and write

$$\nu_k^{(r)} = \frac{\Gamma(k)}{\Gamma(r+k+1)} \frac{(r+k)^{r+k}}{(k-1)^{k-1}}.$$

Then, if $A_r(x) \geq 0$ ($x \geq 0$),

$$(11) \quad \bar{\sigma}_{r+k} \leq \frac{e^{r+1}}{\nu_k^{(r)}} \bar{s},$$

provided the right-hand side is finite.

Proof. From the known result [2, Lemma 6]:

$$(12) \quad A_{r+k}(x) = \frac{\Gamma(r+k+1)}{\Gamma(k)\Gamma(r+1)} \int_0^x (x-u)^{k-1} A_r(u) du,$$

we obtain

$$(13) \quad \sigma_{r+k}(x) = \frac{\Gamma(r+k+1)}{\Gamma(k)\Gamma(r+1)} y^{r+1} \int_0^x g(u) e^{-uy} A_r(u) du,$$

where

$$g(u) = \frac{(x-u)^{k-1} e^{uy}}{x^{r+k} y^{r+1}}$$

reaches its maximum when $u = x - \frac{k-1}{y}$, the maximum being

³ When $k=1$, $(k-1)^{k-1}$ in the expression for $\nu_k^{(r)}$ is to mean 1.

$$\frac{e^{xy}}{(xy)^{k+r}} e^{-(k-1)} (k-1)^{k-1}.$$

Since this maximum is least when $xy = k + r$, (13) gives, with the last-mentioned relation between x and y ,

$$\sigma_{r+k}(x) \leq \frac{e^{r+1}}{v_k(r)} \frac{1}{\Gamma(r+1)} y^{r+1} \int_0^x e^{-uy} du \int_0^u r(u-t)^{r-1} A(t) dt.$$

In this inequality we can invert the order of integration and let $x \rightarrow \infty$ (or $y \rightarrow +0$), obtaining

$$\begin{aligned} \bar{\sigma}_{r+k} &\leq \frac{e^{r+1}}{v_k(r)} \frac{1}{\Gamma(r+1)} \lim_{y \rightarrow +0} y^{r+1} \int_0^\infty A(t) dt \int_t^\infty r(u-t)^{r-1} e^{-uy} du \\ &= \frac{e^{r+1}}{v_k(r)} \frac{1}{\Gamma(r+1)} \lim_{y \rightarrow +0} y \int_0^\infty A(t) e^{-ty} \Gamma(r+1) dt \end{aligned}$$

which is the conclusion sought.

A particular version of Lemma 5, involving the $(r+k)$ -th Cesàro mean and the Abel mean, has been given by Garten and Knopp [1, Satz 4]. The lemma can be used in much the same way as its particular version and made to yield the next two results.

LEMMA 6. *If (in the notation of Lemma 5), for some $r \geq 0$, $\sigma_r(x) \geq -K$, then*

$$\bar{\sigma}_\infty \equiv \lim_{k \rightarrow \infty} \bar{\sigma}_k = \bar{s}.$$

Proof. Starting from the following relation, equivalent to (12),

$$A_{r+k}(x) + Kx^{r+k} = \frac{\Gamma(r+k+1)}{\Gamma(k)\Gamma(r+1)} \int_0^x (x-u)^{k-1} \{A_r(u) + Ku^r\} du,$$

we can get, arguing as in Lemma 5,

$$\bar{\sigma}_{r+k} + K \leq \frac{e^{r+1}}{v_k(r)} (\bar{s} + K)$$

whence, letting $k \rightarrow \infty$, it follows that

$$\bar{\sigma}_\infty \leq \bar{s}$$

which, in conjunction with the universal relation $\bar{s} \leq \bar{\sigma}_\infty$, establishes the conclusion of the theorem.

The above argument tacitly assumes that $\bar{s} < \infty$, the case $\bar{s} = \infty$ being trivial.

⁴ From this point onwards, the proof tacitly assumes $r > 0$, the argument in the case $r = 0$ being an obvious simplification of what follows.

2.1. Lemma 6 can be stated in a more comprehensive form as follows.

THEOREM 2. *Let*

$$\overline{\lim}_{t \rightarrow +0} F(t) = \bar{s}, \quad \overline{\lim}_{x \rightarrow \infty} \sigma_k(x) = \underline{\sigma}_k, \quad \lim_{k \rightarrow \infty} \underline{\sigma}_k = \underline{\sigma}_\infty.$$

Then (a) provided $\underline{\sigma}_\infty$ is finite, and consequently, for large enough r , $\sigma_r(x)$ is bounded below,

$$\bar{\sigma}_\infty = \bar{s};$$

(b) provided $\bar{\sigma}_\infty$ is finite,

$$\underline{\sigma}_\infty = \underline{s};$$

(c) provided $\underline{\sigma}_\infty$ and $\bar{\sigma}_\infty$ are both finite,

$$\underline{\sigma}_\infty = \underline{s}, \quad \bar{\sigma}_\infty = \bar{s}.$$

3. **Concluding remarks.** Suppose that $\bar{\sigma}_\infty$ and $\underline{\sigma}_\infty$ have the common finite value s . Then, for large enough r , $\bar{\sigma}_r$ and $\underline{\sigma}_r$ are finite. And so $\sigma_r(x)$ is bounded for $r \geq k$ (say), which ensures the hypotheses (3) of Theorem 1. Further, the universal fact $\underline{\sigma}_\infty \leq \overline{\lim}_{t \rightarrow +0} F(t) \leq \bar{\sigma}_\infty$ implies the hypothesis (2) of Theorem 1, viz., $\lim_{t \rightarrow +0} F(t) = s$. Therefore, appealing to Theorem 1, we find that $\bar{\sigma}_r = \underline{\sigma}_r = s$ for $r \geq k + 1$. Thus we establish

THEOREM *t.* In the notation of Theorem 2, $\bar{\sigma}_\infty = \underline{\sigma}_\infty = s$ (finite) ensures $\bar{\sigma}_r = \underline{\sigma}_r = s$ for all sufficiently large r .

Next supposing that \bar{s} and \underline{s} have the common finite value s , and further that $\underline{\sigma}_k > -\infty$, we can conclude from Theorem 2 that $\bar{\sigma}_\infty = s = \underline{\sigma}_\infty$. This fact, taken along with Theorem *t*, proves

THEOREM *T.* In the notation of Theorem 2, $\bar{s} = \underline{s} = s$ (finite) and $\underline{\sigma}_k > -\infty$ together ensure $\bar{\sigma}_r = \underline{\sigma}_r = s$ for all sufficiently large r .

The forms of Theorems 2, *t*, *T* involving Cesàro means and the Abel mean are well known [3, Theorems 1(a), (b), (c); 4, Theorems *t*, *T*].

If $A(x) = \sum_{\lambda_n \leq x} a_n$ ($0 < \lambda_n < \lambda_{n+1} \rightarrow \infty$) in §§ 2, 3, the propositions given there assume the form of statements connecting the summability of $\sum a_n$ by Riesz means with the summability of the series by the $D(\lambda_n)$ -process, i. e., by Dirichlet's series.

REFERENCES.

1. V. Garten and K. Knopp, "Ungleichungen zwischen Mittelwerten von Zahlenfolgen und Funktionen," *Mathematische Zeitschrift*, vol. 42 (1937), pp. 365-388.
2. G. H. Hardy and M. Riesz, *The general theory of Dirichlet's series* (Cambridge Tracts, No. 18, 1915).
3. V. Ramaswami, "Some Tauberian theorems on oscillations," *Journal of the London Mathematical Society*, vol. 10 (1935), pp. 294-308.
4. J. E. Littlewood, "Note on the preceding paper," *Journal of the London Mathematical Society*, vol. 10 (1935), pp. 309-310.
5. O. Szász, "Converse theorems of summability for Dirichlet's series," *Transactions of the American Mathematical Society*, vol. 39 (1936), pp. 117-130.
6. J. Karamata, "Neuer Beweis und Verallgemeinerung einiger Tauberian-Sätze," *Mathematische Zeitschrift*, vol. 33 (1931), pp. 294-299.

A STUDY ON THE THEORY OF CONJUGATE NETS.*

By CHUAN-CHIH HSIUNG.

1. Introduction. This paper is concerned with the study of the projective differential geometry of curves of a conjugate net on an analytic non-ruled surface in ordinary space. Some portions of the theory of a surface S referred to a conjugate net N_x , which are used in later developments, are summarized in **2**.

Let C_λ be any curve on the surface S through a general point P_x . As the point P_x moves along the curve C_λ , the parametric u -, v -tangents generate respectively two non-developable ruled surfaces $R^{(u)}$, $R^{(v)}$. In **3** we find the correspondence between the line joining any two points T , \bar{T} on the parametric u -, v -tangents through the point P_x and the intersection of the tangent planes of the ruled surfaces $R^{(u)}$, $R^{(v)}$ at the points T , \bar{T} respectively. In **4** two quadrics are introduced of which one has contact of the second order with the ruled surface $R^{(u)}$ at the point T and contact of the first order with the ruled surface $R^{(v)}$ at the point \bar{T} , and the other has the same properties with the roles of u , v interchanged.

By a *line* l_1 we mean, as usual, any line through the point P_x of the surface S and not lying in the tangent plane of the surface S at P_x ; by a *line* l_2 we mean, dually, any line in the tangent plane of the surface S at P_x but not passing through the point P_x . In **5**, by means of certain correspondences between l_1 and l_2 we construct a one-parameter family of polarities and present a new geometric characterization for a general canonical line (of each kind) of Davis.

In the last section we derive two one-parameter families of lines, of which two may be regarded as generalizations of the canonical edges of Green of the asymptotic net of a surface to a conjugate net.

2. Analytic basis. If the projective homogeneous coordinates $x^{(1)}$, \dots , $x^{(4)}$ of a point P_x in ordinary space are given as analytic functions of two independent variables u , v by equations of the form

$$(1) \quad x = x(u, v),$$

* Received September 19, 1946.

the locus of P_x as u, v vary is an analytic surface \mathcal{S} . If the parametric curves on the surface form a conjugate net N_x , the four coordinates x and the four coordinates of the point P_y , which is the harmonic conjugate of the point P_x with respect to the foci of the axis of the point P_x , satisfy a system of partial differential equations of the form [7, pp. 250-254]

$$(2) \quad \begin{aligned} x_{uu} &= px + \alpha x_u + Ly, \\ x_{uv} &= cx + \alpha x_u + bx_v, \\ x_{vv} &= qx + \delta x_v + Ny \end{aligned} \quad (LN \neq 0),$$

subscripts indicating partial differentiation and the coefficients being functions of u, v which satisfy certain integrability conditions. It is easy to verify that

$$(3) \quad y_u = fx - nx_u + sx_v + Ay, \quad y_v = gx + tx_u + nx_v + By,$$

where we have placed

$$(4) \quad \begin{aligned} fN &= c_v + ac + bq - c\delta - q_u, & gL &= c_u + bc + ap - c\alpha - p_v, \\ -nN &= a_v + a^2 - a\delta - q, & tL &= a_u + ab + c - \alpha_v, \\ sN &= b_v + ab + c - \delta_u, & nL &= b_u + b^2 - b\alpha - p, \\ A &= b - (\log N)_u, & B &= a - (\log L)_v. \end{aligned}$$

The ray-points, or Laplace transformed points, ρ, σ of the point P_x are given by the formulas

$$(5) \quad \rho = x_u - bx, \quad \sigma = x_v - ax.$$

Some of the invariants of the parametric conjugate net are

$$(6) \quad \begin{aligned} H &= c + ab - a_u, & K &= c + ab - b_v, \\ \mathfrak{H} &= sN, & \mathfrak{K} &= tL, \\ 8\mathfrak{B}' &= 4a - 2\delta + l_v, & 8\mathfrak{C}' &= 4b - 2\alpha - l_u, \\ r &= N/L, \end{aligned}$$

where l is defined by placing

$$l = \log r.$$

We shall suppose that $\mathfrak{H}\mathfrak{K} \neq 0$ so that the parametric curves are not plane curves.

If the points x, ρ, σ, y are used as the vertices of a local tetrahedron of reference with a unit point suitably chosen, then any point given by an expression of the form

$$(7) \quad x_1x + x_2\rho + x_3\sigma + x_4y$$

has local coordinates proportional to x_1, \dots, x_4 .

Introducing non-homogeneous coordinates by the definitions

$$(8) \quad x = x_2/x_1, \quad y = x_3/x_1, \quad z = x_4/x_1,$$

we obtain a power series expansion [6, p. 481] for one non-homogeneous coordinate z of a point on the surface S in terms of the other two coordinates x, y :

$$(9) \quad z = \frac{1}{2}(Lx^2 + Ny^2) + \frac{4}{3}(L\mathfrak{L}'x^3 + N\mathfrak{B}'y^3) + \dots$$

3. A Correspondence. Let us consider a point P_x on a surface S and the curve C_λ passing through P_x and belonging to the family defined on S by the equation

$$(10) \quad dv - \lambda du = 0,$$

where λ is a function of u, v ; and let the parametric representation of the curve C_λ be

$$(11) \quad u = u(w), \quad v = v(w).$$

Without loss of generality we may assume that, for the point P_x , $w = 0$. Throughout this paper, differentiation with respect to w is denoted by accents and $u', u'', \dots; v', v'', \dots$ are the derivatives of u, v at any point on the parametric u -tangent through the point P_x , so that

$$(12) \quad \lambda = v'/u', \quad \lambda' = v''/u'^2 - v'u''/u'^3.$$

The parametric u -, v -tangents at points of the curve C_λ generate two non-developable ruled surfaces $R^{(u)}, R^{(v)}$ respectively. The points

$$(13) \quad T = ex - \rho, \quad \bar{T} = \bar{e}x - \sigma,$$

where e, \bar{e} are functions of u, v , lie on the parametric u -, v -tangents through the point P_x respectively, and the line l_2 determined by them lies in the tangent plane of the surface S at the point P_x .

From system (2) and equations (3), by differentiation and substitution, any derivative of T with respect to u, v can be expressed as a linear combination of x, x_u, x_v, y . In particular, making use of equations (4), (6) one obtains

$$\begin{aligned}
 T_u &= (e_u + b_u - p)x + (e + b - \alpha)x_u - Ly, \\
 T_v &= (e_v + b_v - c)x - ax_u + ex_v, \\
 (14) \quad T_{uu} &= (*)x + (*)x_u - sLx_v + [L(e + b - \alpha) - L_u - AL]y, \\
 T_{uv} &= (*)x + (*)x_u + (e_u + eb)x_v - aLy, \\
 T_{vv} &= (*)x + (*)x_u + (2e_v + e\delta - K)x_v + eNy.
 \end{aligned}$$

where $(*)$ denotes terms immaterial for our purpose.

The tangent planes to the ruled surface $R^{(u)}$ at the point T and to the ruled surface $R^{(v)}$ at the point \bar{T} intersect in a line l_1 which passes through the point P_x . There must be a point on the line l_1 given by an expression of the form

$$z = y + \xi\rho + \eta\sigma.$$

The plane of the points x, z, T is to be tangent to the ruled surface $R^{(u)}$ at T , and therefore must contain the tangent to the curve traced by T as the point P_x moves along the curve C_λ . It is clear from the first two of equations (14) that a point on this tangent is given by

$$T_w = (*)x + (*)\rho + ev'\sigma - Lu'y;$$

this is to lie in the plane xzT , which can occur if and only if $\eta = -ev'/Lu'$. Similarly, the plane $xz\bar{T}$ is tangent to the ruled surface $R^{(v)}$ at \bar{T} if and only if $\xi = -\bar{e}u'/Nv'$. Thus we obtain a correspondence between the line l_2 in the tangent plane $x_4 = 0$ and the line l_1 through the point P_x :

$$(15) \quad Lx_3 + e\lambda x_4 = 0, \quad N\lambda x_2 + \bar{e}x_4 = 0.$$

It is easily seen that *this correspondence becomes a polarity with respect to a quadric (and therefore ∞^1 quadrics) if and only if λ satisfies*

$$(16) \quad L - N\lambda^2 = 0,$$

that is, *if and only if the tangent of the curve C_λ at the point P_x is an associate conjugate tangent of the parametric conjugate net N_x* . The equation of any one of these quadrics is found to be of the form

$$(17) \quad \sqrt{LN} x_2 x_3 - x_1 x_4 + k_4 x_4^2 = 0,$$

where k_4 denotes a parameter. If a unique quadric of this pencil is desired, we may choose the one that passes through the covariant point P_y . For this quadric we have $k_4 = 0$.

4. Two quadrics associated with the points T, \bar{T} . The parametric vector equation of the ruled surface $R^{(u)}$ is

$$(18) \quad X = T + \mu x,$$

where μ, w are taken as independent parameters. The coordinates of any point \mathfrak{X} near the point T on the ruled surface $R^{(u)}$ can be represented by Taylor's expansion as power series in the increments $\Delta\mu, \Delta w$ corresponding to displacement on $R^{(u)}$ from T to the point \mathfrak{X} :

$$(19) \quad \mathfrak{X} = X(\Delta\mu, \Delta w) = X(0, 0) + X_\mu(0, 0)\Delta\mu + X_w(0, 0)\Delta w \\ + \frac{1}{2}[X_{\mu\mu}(0, 0)\Delta\mu^2 + 2X_{\mu w}(0, 0)\Delta\mu\Delta w + X_{ww}(0, 0)\Delta w^2] + \dots,$$

where

$$(20) \quad \begin{aligned} X_\mu(0, 0) &= x, & X_w(0, 0) &= T_u u' + T_v v', \\ X_{\mu\mu}(0, 0) &= 0, & X_{\mu w}(0, 0) &= x_u u' + x_v v', \\ X_{ww}(0, 0) &= T_{uu} u'^2 + 2T_{uv} u'v' + T_{vv} v'^2 + T_{uu} u'' + T_{vv} v''. \end{aligned}$$

In accordance with equation (14), all the derivatives of X at the point T may be expressed uniquely as linear combinations of x, ρ, σ, y , so that we may write \mathfrak{X} in the form (7), where the local coordinates x_1, \dots, x_4 of the point \mathfrak{X} are given by the expansions

$$(21) \quad \begin{aligned} x_1 &= e + \Delta\mu + [u'(e_u + eb + nL) + v'(e_v + ea - K)]\Delta w + \dots, \\ x_2 &= -1 + [u'(e + b - \alpha) - av']\Delta w + \dots, \\ x_3 &= ev'\Delta w + v'\Delta\mu\Delta w + \frac{1}{2}[-sLu'^2 + 2(e_u + eb)u'v' \\ &\quad + (2e_v + e\delta - K)v'^2 + ev'']\Delta w^2 + \dots, \\ x_4 &= -Lu'\Delta w + \frac{1}{2}\{[L(e + b - \alpha) - L_u - AL]u'^2 \\ &\quad - 2aLu'v' + eNv'^2 - Lu''\}\Delta w^2 + \dots. \end{aligned}$$

The equation of any quadric, which passes through the points T, \bar{T} and is tangent to the planes (15) at T, \bar{T} respectively, is found to be

$$(22) \quad a_{11}(x_1^2 + 2ex_1x_2 + 2\bar{e}x_1x_3 + e^2x_2^2 + \bar{e}^2x_3^2) + 2a_{14}x_1x_4 + 2a_{23}x_2x_3 \\ + 2a_{24}x_2x_4 + 2a_{34}x_3x_4 + a_{44}x_4^2 = 0,$$

where the coefficients a_{24}, a_{34} are defined by the following formulas:

$$(23) \quad \begin{aligned} a_{24} &= (d/Lu')(Lu'a_{14} + v'a_{23} - e\bar{e}v'a_{11}), \\ a_{34} &= (e/Nv')(Nv'a_{14} + u'a_{23} - e\bar{e}u'a_{11}). \end{aligned}$$

Imposing on equation (22) the further conditions that it be satisfied by the series (21) identically in the terms of the second degree of $\Delta\mu$, Δv , we obtain easily the equation of the quadric Q_u which has contact of the second order with the ruled surface $R^{(u)}$ at the point T and contact of the first order with the ruled surface $R^{(v)}$ at the point \bar{T} , namely,

$$(24) \quad 2L(\lambda x_1 x_4 - L x_2 x_3) - (2\bar{e}/r\lambda)(L - N\lambda^2)x_3 x_4 + \{L(2n\lambda - \lambda^2 K/L \\ + \mathfrak{S}/N) + e\lambda[4(\mathfrak{C}' + \lambda\mathfrak{B}') - \frac{1}{2}(l_u + \lambda l_v) - (\log \lambda)'] \\ - e(2\bar{e}/N - e\lambda/L)(L - N\lambda^2)\}x_4^2 = 0.$$

There is also a unique quadric Q_v with second order contact with the ruled surface $R^{(v)}$ at the point \bar{T} and first order contact with the ruled surface $R^{(u)}$ at the point T . The equation of this quadric Q_v can be obtained in a way similar to the foregoing, or else can be written immediately by applying the substitution

$$(25) \quad \begin{pmatrix} u & 2 & e & s & n & L & H & \mathfrak{B}' & \lambda & \lambda' \\ v & 3 & \bar{e} & t & -n & N & K & \mathfrak{C}' & 1/\lambda & -\lambda'/\lambda^3 \end{pmatrix}$$

to equation (24); the result is

$$(26) \quad 2N(x_1 x_4 - N\lambda x_2 x_3) + 2er(L - N\lambda^2)x_2 x_4 + \{-N(2n + H/\lambda N \\ - \lambda\mathfrak{B}'/L) + (\bar{e}/\lambda)[4(\mathfrak{C}' + \lambda\mathfrak{B}') + \frac{1}{2}(l_u + \lambda l_v) + \log \lambda'] \\ + (\bar{e}/\lambda^2)(2e\lambda/L - \bar{e}/N)(L - N\lambda^2)\}x_4^2 = 0.$$

On the other hand, if π be any plane passing through the line $T\bar{T}$, then the plane sections Γ_u , Γ_v of the ruled surfaces $R^{(u)}$, $R^{(v)}$ made by π have simple points at T , \bar{T} respectively. It should be noted¹ that as the plane π revolves about the line $T\bar{T}$, the locus of the conic in π having contact of the second order with the plane section Γ_u at the point T and contact of the first order with the plane section Γ_v at the point \bar{T} is the quadric Q_u , and a similar argument can be made with u , v interchanged.

Since the quadrics Q_u , Q_v both depend upon the points T , \bar{T} , we obtain two two-parameter families of quadrics associated with the point P_x of the curve C_λ . In particular, if the points T , \bar{T} are the ray-points ρ , σ of the point P_x , then $e = \bar{e} = 0$, and the equations of the quadrics Q_u , Q_v become respectively

$$(27) \quad 2(\lambda x_1 x_4 - L x_2 x_3) + (2n\lambda - \lambda^2 K/L + \mathfrak{S}/N)x_4^2 = 0,$$

¹ For the proof of this fact for two hypersurfaces and its generalization see paper [4] of the author.

$$(28) \quad 2(x_1x_4 - N\lambda x_2x_3) - (2n + H/\lambda N - \lambda\mathfrak{R}/L)x_4^2 = 0.$$

The two quadrics (27), (28) intersect, not only in the parametric tangents, but also in a residual conic whose plane is obtained by eliminating the terms containing x_2x_3 from equations (27), (28), namely,

$$(29) \quad 2(L - N\lambda^2)x_1 + [rK\lambda^3 + \lambda(\mathfrak{R} - \mathfrak{S}) - H/\lambda r - 2n(L + N\lambda^2)]x_4 = 0,$$

which is a plane passing through the ray $\rho\sigma$. Moreover, the cone projecting this conic of intersection from the point P_x is similarly found to be given by the equation

$$(30) \quad 2(L - N\lambda^2)x_2x_3 - [4n\lambda + (1/N)(H + \mathfrak{S}) - (\lambda^2/L)(K + \mathfrak{R})]x_4^2 = 0.$$

5. The canonical lines of Davis. A canonical line $l_1(h)$ of the first kind of Davis [1, p. 18] is the intersection of the two planes

$$(31) \quad Lx + h\mathfrak{C}'z = 0, \quad Ny + h\mathfrak{B}'z = 0,$$

and a canonical line $l_2(h)$ of the second kind crosses the parametric tangents at the point P_x in the points

$$(32) \quad (-1/h\mathfrak{C}'), 0, 0) \quad (0, -(1/h\mathfrak{B}'), 0),$$

where h is a constant.

In order to give a new geometric characterization of a general canonical line of each kind of Davis we derive a one-parameter family of correspondences as follows.

Let \bar{g} be the polar line of any line l_1 with the equation

$$(33) \quad x + ez = 0, \quad y + \bar{e}z = 0,$$

with respect to any quadric of Darboux,

$$(34) \quad Lx^2 + Ny^2 + z(-2 + 4\mathfrak{C}'x + 4\mathfrak{B}'y + k_4z) = 0,$$

at the point P_x of the surface S , where k_4 is a parameter. If \bar{g} intersects the u -, v -tangents respectively in the points $P_1^{(\infty)}(\bar{x}^{(\infty)}, 0, 0)$ and $\bar{P}_2^{(\infty)}(0, \bar{y}^{(\infty)}, 0)$, then

$$(35) \quad 1/\bar{x}^{(\infty)} = 2\mathfrak{C}' - eL, \quad 1/\bar{y}^{(\infty)} = 2\mathfrak{B}' - \bar{e}N.$$

This correspondence between l_1 and the line joining the points $\bar{P}_1^{(\infty)}$, $\bar{P}_2^{(\infty)}$ is called, for brevity, the correspondence C_∞ .

On the other hand, let Γ_1 be the plane section of the surface S made by the plane through l_1 and the u -tangent; then by means of (9) we obtain easily the power series expansion of Γ_1 at P_x :

$$(36) \quad y = -(\bar{e}/2)Lx^2 - \frac{4}{3}\bar{e}L\mathfrak{C}'x^3 + \dots$$

The pole $\bar{P}_1^{(0)}(\bar{x}^{(0)}, 0, 0)$ of l_1 with respect to any four-point conic of Γ_1 at P_x is given by

$$(37) \quad 1/\bar{x}^{(0)} = \frac{2}{3}\mathfrak{C}' - eL.$$

For the plane section Γ_2 of the surface S made by the plane through l_1 and the v -tangent, we obtain similarly on the v -tangent the pole $\bar{P}_2^{(0)}(0, \bar{y}^{(0)}, 0)$, where

$$(38) \quad 1/\bar{y}^{(0)} = \frac{2}{3}\mathfrak{B}' - \bar{e}N.$$

Thus we arrive at a correspondence C_0 between l_1 and the line joining the points $\bar{P}_1^{(0)}, \bar{P}_2^{(0)}$.

Now on the u -, v -tangents we take respectively two points $\bar{P}_1^{(k)}, \bar{P}_2^{(k)}$ such that the cross ratio of the four points $\bar{P}_m^{(\infty)}, \bar{P}_m^{(0)}, P_x$ and $\bar{P}_m^{(k)}$,

$$(39) \quad (\bar{P}_m^{(\infty)}\bar{P}_m^{(0)}, P_x\bar{P}_m^{(k)}) = k \quad (m = 1, 2),$$

k being a constant, and define a one-parameter family of correspondences C_k between l_1 and the join of $\bar{P}_1^{(k)}, \bar{P}_2^{(k)}$ by the equation

$$(40) \quad z = \left[\frac{2}{3}\left(\frac{4-3k}{1-k}\right)\mathfrak{C}' - eL\right]x + \left[\frac{2}{3}\left(\frac{4-3k}{1-k}\right)\mathfrak{B}' - \bar{e}N\right]y - 1 = 0.$$

It is easily shown that the correspondence C_k is the polarity with respect to any quadric Q_k of the pencil,

$$(41) \quad Lx^2 + Ny^2 + z[-2 + \frac{4}{3}\left(\frac{4-3k}{1-k}\right)\mathfrak{C}'x + \frac{4}{3}\left(\frac{4-3k}{1-k}\right)\mathfrak{B}'y + k_1z] = k_2$$

k_1 being a parameter.^{1a}

^{1a} By means of the Fubini canonical differential equations of a surface S referred to its asymptotic net we can prove that the pencil (41) of quadrics associated with the tangents of the curves of a conjugate net at an ordinary point P_x is in the family of conjugal quadrics of S at P_x (V. G. Grove, "The transformation of Čech," *Bulletin of the American Mathematical Society*, vol. 50 (1944), pp. 231-234). Furthermore, if we define the two cross ratios in equations (39) respectively by two different constants k_1 and k_2 instead of k , we may obtain a family of quadrics, which is an extension of the pencil (41) but belongs to the more general family defined by Grove ("Quadrics associated with a curve on a surface," *Bulletin of the American Mathematical Society*, vol. 51 (1945), pp. 281-287).

Observing that any quadric having contact of the second order with the surface S at the point P_x is given by

$$(42) \quad Lx^2 + Ny^2 + z[-2 + k_2x + k_3y + k_4z] = 0,$$

k_2, k_3, k_4 being arbitrary parameters, we infer immediately that the quadrics of the pencil (41) belong to this family.

If $k = \frac{4}{3}$ and Q_k passes through the point P_y , we have the canonical quadric of Davis [1, p. 10]; if $k = \infty$, we have the quadrics of Darboux; and if $k = 0$, we have the pencil [8, p. 697] of the quadrics having contact of the third order with the parametric curves at the point P_x of the surface S . Thus we have proved the

THEOREM. *The correspondence C_0 is the polarity with respect to the pencil of the quadrics having contact of the third order with the parametric curves at the point P_x of the surface S .*

Since the polar line of the axis with respect to the quadric Q_k is a general canonical line of the second kind of Davis,

$$(43) \quad z = \frac{2}{3} \left(\frac{4-3k}{1-k} \right) \mathfrak{C}'x + \frac{2}{3} \left(\frac{4-3k}{1-k} \right) \mathfrak{B}'y - 1 = 0,$$

and the polar line of the ray with respect to the quadric Q_k is a general canonical line of the first kind of Davis,

$$(44) \quad Lx + \frac{2}{3} \left(\frac{4-3k}{1-k} \right) \mathfrak{C}'z = 0, \quad Ny + \frac{2}{3} \left(\frac{4-3k}{1-k} \right) \mathfrak{B}'z = 0,$$

we obtain the ²

THEOREM. *The polar lines of the axis and the ray with respect to the quadric Q_k are respectively the canonical lines (43) and (44) of the second and the first kinds of Davis which depend upon the quadric Q_k .*

By a proper selection of the constant k these lines may be made to become any desired canonical lines of both kinds of Davis. In particular, when $k = \infty$, the lines (43) and (44) are respectively the associate ray and the associate axis. When $k = 0$, we obtain the principal join ³ and the canonical line $l_1(\frac{8}{3})$ of the first kind.⁴

² A general canonical line of each kind of the surface referred to the asymptotic curves has been characterized geometrically in a similar manner. See my paper [3].

³ For the definition of the principal join see E. P. Lane [5, p. 702].

⁴ These two characterizations were given by Lane and MacQueen [8, pp. 698-699].

Moreover, the polar line of any canonical line $l_1(k_1)$ of the first kind with respect to the quadric Q_k is a canonical line $l_2(k_2)$ of the second kind which depends on the selection of the constant k ; and the values of k_1 and k_2 are the same when and only when $k = \frac{4}{3}$, that is, when and only when Q_k is a quadric in the canonical pencil of Davis.⁵

6. Two one-parameter families of lines. The equation of the family (41) of quadrics may be rewritten in homogeneous coordinates as

$$(45) \quad Lx_2^2 + Nx_3^2 + x_4(-2x_1 + k\mathfrak{E}'x_2 + k\mathfrak{B}'x_3 + k_4x_4) = 0,$$

where k is a new constant. The polar line of the line $l_1(33)$ with respect to a general quadric of the family (45) is found to intersect the u -, v -tangents through the point P_x respectively in the points

$$(46) \quad \Phi_k = (k\mathfrak{E}' - 2eL)x + 2\rho, \quad \Psi_k = (k\mathfrak{B}' - 2eN)x + 2\sigma.$$

On the other hand, the homogeneous coordinates of a point X near P_x and on the u -curve through P_x can be expressed as power series in the increment Δu corresponding to displacement from P_x to the point X along the u -curve, namely [8, p. 693],

$$(47) \quad \begin{aligned} x_1 &= 1 + b\Delta u + \cdots, \\ x_2 &= \Delta u + \frac{1}{2}\alpha\Delta u^2 + \cdots, \\ x_3 &= \frac{1}{6}(\mathfrak{E}/r)\Delta u^3 + \cdots, \\ x_4 &= \frac{1}{2}L\Delta u^2 + \frac{1}{6}L(\alpha + b - l_u)\Delta u^3 + \cdots. \end{aligned}$$

For the purpose of finding the projection of the u -curve onto its osculating plane from any point τ on the line l_1 :

$$(48) \quad \tau = \lambda x - e\rho - e\sigma + y,$$

λ being an arbitrary scalar function of u , v , and in connection with the later development, it is convenient to introduce another local tetrahedron with vertices at the points x , Φ_k , τ , y . If the point in space with coordinates x_1, \dots, x_4 referred to the tetrahedron x, ρ, σ, y has coordinates X_1, \dots, X_4 referred to the new tetrahedron x, Φ_k, τ, y , then the identity

$$x_1x + x_2\rho + x_3\sigma + x_4y = X_1x + X_2\Phi_k + X_3\tau + X_4y$$

⁵ By means of this property Rasmussen and Hagen have defined the quadrics of the pencil (41), but they have made an error in considering k_1 and k_2 as always the same. See their joint paper [9, p. 301].

yields the equations for the transformation of coordinates between the two tetrahedra; after solution for X_1, \dots, X_4 these equations can be written in the form

$$(49) \quad \begin{aligned} X_1 &= x_1 + (eL - \tfrac{1}{2}k\mathfrak{C}')x_2 + 1/\bar{e}[e(\tfrac{1}{2}k\mathfrak{C}' - eL) + \lambda]x_3, \\ X_2 &= \tfrac{1}{2}x_2 - (e/2\bar{e})x_3, & X_3 &= -(1/\bar{e})x_3, \\ X_4 &= (1/e)x_3 + x_4. \end{aligned}$$

The parametric equations of the projection C'_u of the u -curve from the new vertex $(0, 0, 1, 0)$ onto the tangent plane, $X_3 = 0$, are found by substituting the series (47) for x_1, \dots, x_4 into equations (49) and taking such a linear combination of the resulting coordinates X_1, \dots, X_4 and of $0, 0, 1, 0$ as will make the third coordinate vanish. These parametric equations, to terms of as high degree as will be needed, are

$$(50) \quad \begin{aligned} X_1 &= 1 + (b + eL - \tfrac{1}{2}k\mathfrak{C}')\Delta u + \dots, \\ X_2 &= \tfrac{1}{2}\Delta u + \tfrac{1}{4}\alpha\Delta u^2 + \dots, \\ X_4 &= \tfrac{1}{2}L\Delta u^2 + \tfrac{1}{8}L(\alpha + b - l_u + \mathfrak{S}/\bar{e}N)\Delta u^3 + \dots, \end{aligned}$$

the third coordinate being zero in the remainder of this section. Imposing on the general equation of a conic the conditions that it be satisfied by the series (50) for X_1, X_2, X_4 identically in Δu as far as the terms in Δu^3 , we obtain the equation of the four-point conics at the point P_x of the projection C'_u of the u -curve, namely,

$$(51) \quad X_1X_4 - 2LX_2^2 - 2[eL + \mathfrak{S}/3\bar{e}N + (16 - 3k)\mathfrak{C}'/6]X_2X_4 + h_4X_4^2 = 0,$$

where h_4 is a parameter. The polar line of the point $\Phi_k(0, 1, 0)$ with respect to any one of these conics has the equation

$$(52) \quad 2LX_2 + [eL + \mathfrak{S}/3\bar{e}N + (16 - 3k)\mathfrak{C}'/6]X_4 = 0.$$

Since the plane of the line $l_1(33)$ and the v -tangent intersects the osculating plane of the u -curve at the point P_x in the line

$$(53) \quad X_3 = 0, \quad 2X_2 + eX_4 = 0,$$

we know that the line (52) coincides with the line (53) in case

$$(54) \quad \bar{e} = 2\mathfrak{S}/(3k - 16)N\mathfrak{C}'.$$

Similarly, with the roles of the curves u, v interchanged, we get

$$(55) \quad e = 2\mathfrak{R}/(3k - 16)L\mathfrak{B}'.$$

Thus for each pencil of quadrics of the family (45) we can determine two lines, namely, the line $l_1(33)$ satisfying the conditions (54), (55) and its polar line with respect to the pencil of quadrics, and hence *associated with the point P_x of the parametric conjugate net N_x there exist two one-parameter families of lines*, which will be called, respectively, *the first and the second families*. In particular, a reference to the original definition of Green [2, p. 114] shows that when $k = 4$ the two lines of these families may be regarded as generalizations of the canonical edges of Green of the asymptotic net of a surface to a conjugate net. Further, we observe that *all lines of the first family lie in a plane*, whose equation is found to be

$$(56) \quad \S \mathcal{B}'x_2 - r\mathcal{R}'\mathcal{C}'x_3 = 0.$$

Finally, noticing equations (31), (32) it is easy to conclude that *if one of the two families of lines mentioned above is a canonical pencil of Davis, then the other is also, and necessary and sufficient conditions for this are*

$$(57) \quad \S = \mathcal{R} = m\mathcal{B}'\mathcal{C}' \quad (m = \text{const.} \neq 0).$$

MICHIGAN STATE COLLEGE AND
NATIONAL UNIVERSITY OF CHEKIANG.

BIBLIOGRAPHY.

1. W. M. Davis, "Contributions to the theory of conjugate nets," *Chicago doctoral dissertation* (1932).
2. G. M. Green, "Memoir on the general theory of surfaces and rectilinear congruences," *Transactions of the American Mathematical Society*, vol. 20 (1919), pp. 79-153.
3. C. C. Hsiung, "The canonical lines," *Duke Mathematical Journal*, vol. 8 (1941), pp. 738-742.
4. C. C. Hsiung, "The contact of conics and quadrics with two hypersurfaces in space of n dimensions," to appear elsewhere.
5. E. P. Lane, "Invariants of intersection of two curves on a surface," *American Journal of Mathematics*, vol. 54 (1932), pp. 699-706.
6. E. P. Lane, "A canonical power series expansion for a surface," *Transactions of the American Mathematical Society*, vol. 37 (1935), pp. 463-482.
7. E. P. Lane, *A Treatise on Projective Differential Geometry*, The University of Chicago Press, 1942.
8. E. P. Lane and M. L. MacQueen, "The curves of a conjugate net," *Duke Mathematical Journal*, vol. 5 (1939), pp. 692-704.
9. R. B. Rasmussen and B. L. Hagen, "Comments on canonical lines," *Bulletin of the American Mathematical Society*, vol. 47 (1941), pp. 298-302.

SIX-RINGS IN MINIMAL FIVE-COLOR MAPS.*

ARTHUR BERNHART.

Introduction. Kempe¹ and Heawood² have shown that five colors are sufficient for coloring any map on a sphere. We conjecture the existence of some maps for which five colors are necessary, and seek the simplest example. For this end *we here consider only minimal maps*, namely those five-color maps such that any map with fewer regions is four-colorable. This paper investigates the structure of minimal maps by a systematic analysis of rings.

By a *proper n -ring* we mean [1] a cycle of n distinct regions, each adjacent to the regions which precede and succeed it in the cyclic order, [2] but to no other region in the cycle, and [3] dividing the rest of the map into two non-empty sides. This definition is equivalent [in minimal maps] to the usage of Birkhoff³ who introduced the terminology "ring of n regions." The first condition conveys the generic meaning of the word ring as used by many four-color writers without a formal definition. The second condition originated with Birkhoff, and corresponds to what Veblen⁴ intended by a *simple* circuit. For $n = 1, 2, 3$ it adds nothing to the generic meaning, but Birkhoff excluded these cases thereby implying the third condition. A ring divides the regions of a map into three mutually exclusive parts: the regions R of the ring itself, the regions I inside the ring, and the regions O outside the ring. Ordinarily the terms "inside" and "outside" are interchangeable, but whenever the structure of one side is simpler or more fully known we shall refer to it as the inside, achieving thereby an economy of description. In proper rings, I is a proper subset of the map. Whenever we use the term ring in the generic sense with a meaning other than that set forth in the foregoing definition, we shall warn the reader by calling it an *improper* ring. Thus a single region forms an improper 1-ring, two adjacent regions form an improper 2-ring, the regions meeting at a vertex form an improper 3-ring, and the regions participating in an edge form an improper 4-ring. Conversely, these examples of improper n -rings are the only possibilities for

* Received August 8, 1946.

¹ A. B. Kempe, *American Journal of Mathematics*, vol. 2 (1879), pp. 193-200.

² P. J. Heawood, *Quarterly Journal of Mathematics*, vol. 24 (1890), pp. 332-338.

³ G. D. Birkhoff, *American Journal of Mathematics*, vol. 35 (1913), pp. 115-128.

⁴ O. Veblen, *Annals of Mathematics*, vol. 14 (1912-13), pp. 86-94.

$n = 1, 2, 3, 4$ respectively. In this paper unless the context indicates otherwise, all maps are minimal, and all rings are proper.

The most significant result is the 6-ring theorem which lists the colorability of each side of a 6-ring giving all solutions compatible with the Kempe chain reductions. Prior to the main theorem the case of n -rings with $n < 5$ is solved again, but in addition to the familiar result that such rings are necessarily improper, a new conjugate-edge theorem is demonstrated. Birkhoff's unique solution for 5-rings is obtained algebraically by solving a simultaneous set of primary and contingent options. This proof serves as a model for subsequent theorems, whose proofs are more involved but embody the same argumentative procedure. A compact notation is developed such that one "color matrix" of sixteen elements expresses 96 contingent options of the Kempe type.

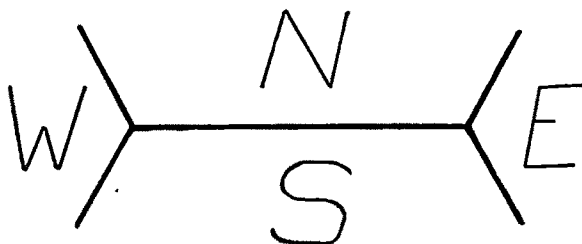
In order to make the discussion of 6-rings more readable, the step-by-step proof of the main theorem has been relegated to a separate section at the end. The significance of the main theorem is discussed. Whereas four-color investigation usually begins with geometric structures and derives their colorability, the methods used here permit arguing from colorability to structure. This innovation proves fruitful, and the main theorem is used to obtain answers to seven structural questions. Each of these results may be read as a theorem in itself revealing new reducible configurations. The Birkhoff result that "a ring of six surrounding four regions is reducible" is extended to a 6-ring surrounding n regions, $3 < n < n_0$ where n_0 is at least 16. Finally, if a 6-ring surrounds j inside regions ($j > 3$), composed of p peripheral regions making contact with the ring plus h additional regions ($j = p + h$), then there are less than $j - 5$ inside pentagons, the p regions form a p -ring with $p > 8$, and $\Sigma(n - 5) = h > 6$, where the summation extends over the j inside n -gons.

4-rings.

Of the four regions around an edge the two which are adjacent will be designated as the *primary components*, the other two the secondary or *guard components*. Let regions N and S be adjacent guarded by regions E and W . If the inside structure of this 4-ring is rearranged to make E and W adjacent thereby separating N and S , the primary and secondary components of the edge are interchanged, and the new edge is *conjugate* to the old.

As a non-trivial corollary to the study of improper 4-rings we state the result for any minimal map: *If one edge is replaced by its conjugate, the*

new map can be 4-colored. This result shows that each map is on the *verge*



of colorability and the slight modification necessary to permit 4-coloring can be achieved *without reducing the number of regions*.

Though the proof is simple we include it here, since the procedure furnishes a prototype for the cases $n > 4$. The entire map cannot be 4-colored, but the outside regions may or may not be colorable consistent with a specified *color scheme* for the ring. We use 1 to indicate one color, 2 a color different from 1, 3 a color different from 1 and 2, 4 a color different from 1, 2 and 3, listing the colors for *SENW* respectively. A 4-ring can be colored in only four schemes, scheme *A*: 1212, scheme *B*: 1213, scheme *C*: 1232, scheme *D*: 1234. On coloring *R* according to any scheme *S* we use the superscript shorthand S^1 inside colorable, S^2 outside colorable, S^3 outside not colorable, S^4 inside not colorable, S^5 neither side colorable, and S^0 both sides colorable. We use a *dot* or mere *juxtaposition* for the logical product (both-and) of two propositions, and a *comma* for the logical option (and/or). When an argument is equally valid for inside and outside the superscripts may be omitted. Thus (S^1, S^4) and (S^2, S^3) are logical options required by the law of excluded middle, while S^1S^4 and S^2S^3 are contradictions; $S^3S^4 = S^5$ and $S^1S^2 = S^0$ but S^0 is absurd since by definition minimal maps are not four-colorable.

Since the primary components *N* and *S* are adjacent, the edge itself can be colored only by schemes *C* and *D*. In the shorthand notation $(CD)^1(AB)^4$. Now if C^2 , then $C^1C^2 = C^0$, which is absurd. Therefore C^3 . Likewise D^3 . But deleting the edge between *N* and *S* without disturbing the outside structure forms a map with fewer than the minimal number of regions, and hence this modified map can be 4-colored, therefore (A^2, B^2) . We refer to this as a *primary or reducing option*. But Kempe¹ has pioneered the way in showing that if the side of an n -ring is colorable according to one color scheme (on the bounding ring) then it must be colorable also according to certain other

schemes. Thus $A \rightarrow (B, C)$; $B \rightarrow (A, D)$; $C \rightarrow (A, D)$; and $D \rightarrow (B, C)$. We refer to such implications as secondary or *contingent options*. The option (B, C) may be satisfied by any one of three alternatives: B and not- C , C and not- B , both B and C .

Kempe derived the contingent option $D : 1234 \rightarrow (B : 1214, C : 1232)$ by introducing the notion of chains. Regions colored 1 and 3 form an "odd" league, while regions colored 2 and 4 form an "even" league. If—under scheme D —regions $S : 1$ and $N : 3$ can be connected by an uninterrupted chain of outside regions belonging to the odd league, then this chain partitions the outside and its bounding ring into two subdivisions. (The chain regions may be assigned arbitrarily to either subdivision.) In one subdivision Kempe retained the original colors, but in the other subdivision he interchanged colors 2 and 4. This exhibits a new 4-coloring of the outside subject to a correspondingly new color scheme on the ring, in this instance scheme C .

But if no such odd-chain connecting S and N exists, Kempe argued that E and W must be connected by an even chain. Interchanging 1 and 3 in one partition formed by this chain, he obtained scheme B , for 1214 and 1213 are dual notations for the same scheme.

Besides the leagues 1-3 versus 2-4 we could use 1-2 vs 3-4 or 1-4 vs. 2-3. The same topological argument is applicable and can be extended to rings $n > 4$, with precisely similar results for schemes with odd-even-odd-even ring coloring. If the ring coloring degenerates to a single odd-even pattern, as in $B : 1213$, the argument is sterile, but in the general case three contingent options are obtained by chain deduction from each postulated scheme.

In the improper 4-ring case under consideration, the four contingent options can be diagrammed by putting the schemes at the corners of a square, A and D placed diagonally. Then each scheme implies an adjacent corner. Since $C^3 D^3$, from the option (A^2, B^2) we obtain (A^2, B^2) and in particular B^2 . This establishes our conjugate-edge theorem.

If the 4-ring were proper we would have four primary options $(A, B)(A, C)(C, D)(B, D)$ for both sides, which can be satisfied only by putting A, D and B, C on opposite sides. But this is contrary to the contingent options, hence proper 4-rings are excluded from minimal maps.

The exclusion of proper 1-rings, 2-rings, and 3-rings is simpler since it does not require chain deductions. For in each case only one color scheme is possible for the ring. Primary options compel the coloring of each side by at least one scheme, and with both sides colorable with the same scheme, the map would be 4-colorable.

5-rings.

Let regions A, B, C, D, E constitute a proper 5-ring. Only ten schemes are possible: scheme A with $ABCDE : 12323$ respectively; scheme A^* with $ABCDE : 12342$; scheme B with $BCDEA : 12323$; etc. The notation emphasizes the cyclic symmetry. Scheme A uses only three colors in which the color of region A is not repeated. Scheme A^* uses four colors in which region A lies between two regions colored the same.

We have the chain deduction $A : 12323 \rightarrow (E : 12123, D^* : 12343)$, and by clock symmetry $A \rightarrow (B, C^*)$. Again $A^* : 12342 \rightarrow (D : 12142, C^* : 12324)$ and by clock symmetry $A^* \rightarrow (C, D^*)$. Cyclic symmetry gives us twenty contingent options, two for each scheme.

Merging the inside with regions B and D we form a modified map with fewer than the minimal number of regions, and which therefore can be 4-colored. This yields the option (A^*, C, D) . Merging region A with the inside yields (A, A^*, C^*, D^*) . Symmetry provides a total of ten primary options, applicable to both outside and inside.

We seek all solutions for this system of 10 primary and 20 secondary options. Trying $(C, D)^2$ we are led directly to the solution $(A, B, C, D, E)^2 (A^*, B^*, C^*, D^*, E^*)^1$. In trying for other solutions we must avoid putting two "consecutive unstarred" schemes on the same side. But then the try $(A)^2$ implies $(A, C^*, D^*)^2 (A^*)^1$ hence $(A^*, C, D)^1$ contrary to the primary option $(A^*, C, D)^2$. Avoiding $(A)^2$ and by symmetry also $(B)^2, (A)^1$ etc., we would have to exclude A, B, C, D, E from both sides. But then the option (A^*, C, D) would compel A^* for both sides, contrary to the definition of minimal maps. This shows that the given solution is the only one. It corresponds to an inside composed of a single pentagon.

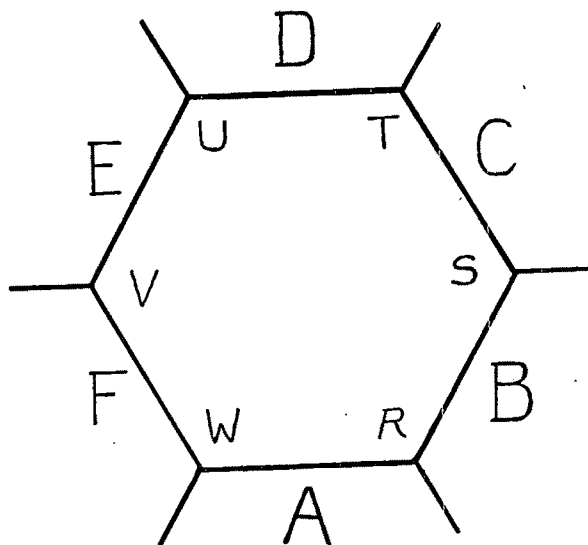
For if the inside contained more than one region, a merging of the inside to form a single region would provide the outside option $(A, B, C, D, E)^2$ leading to the solution $(A, B, C, D, E)^2$. But then the corresponding inside option $(A, B, C, D, E)^1$ could not be satisfied. Thus the case of an inside composed of more than one region is excluded. Hence the inside of a 5-ring must be a single region, a pentagon, and the outside must be colorable by each of the schemes using 4 colors on the ring, and by none other. It is to be noted that minimal maps must include 5-rings, one for each pentagon. By contrast the existence of 6-rings is only conjectural.

6-rings.

Let regions A, B, C, D, E, F constitute a proper 6-ring. Thirty-one schemes are possible:

121212 = X	123123 = 0	123242 = H
213 = F	124 = 9	243 = C^*
232 = E	132 = 1	412 = T
234 = V	134 = D^*	413 = F^*
312 = D	142 = A^*	414 = S
313 = B	143 = 8	423 = 7
314 = G	212 = C	424 = E^*
323 = 2	213 = 3	432 = 4
324 = B^*	214 = 6	434 = R
342 = U	232 = A	
343 = 5	234 = W	

The notation emphasizes the following cyclic groups $(ABCDEF)$ $(A^*B^*C^*D^*E^*F^*)$ $(RSTUVW)$ (GH) (X) (0) (123) (456) (789) .



We now derive 38 primary options, each applicable to both sides. Merging region A with inside: $(A, C^*, E^*, H, R, W, 4, 7)$, six variants. Merging regions A and D with inside: $(A^*, D^*, 0, 1, 8, 9)$, three variants. Merging regions B, D, F and inside: (A, C, E, H, X) , two variants. Merging regions C, E and inside: $(A, D^*, R, W, 1, 4)$, six variants. Merging B and F , also C and E through inside: $(A, D, X, 1, 4)$, three variants. Merging B and F , also making C and E adjacent, through inside: (A^*, C, E, H, T, U) , six variants. Partitioning inside into two regions by an edge guarded by A

and D , then merging these subdivisions of the inside with B and E respectively: $(A^*, D^*, R, U, 1, 4, 5, 8)$, three variants, and clock symmetric: $(A^*, D^*, T, W, 1, 4, 6, 9)$, three variants. Finally, partitioning inside into three pentagons by edges guarded by B, D, F then annexing two pentagons to C and E , respectively: $(B^*, C, E, F^*, T, U, 0, 2, 3, 7)$, six variants. The modified map has fewer regions except when the inside consists of a single hexagon. For this exceptional case the option is justified for the outside by conjugating the edge guarded by C and E .

Merging the inside into a single hexagon yields $(X, A, B, C, D, E, F, 0, 1, 2, 3)$, an option not applicable to the outside of a 6 R 1. Making two pentagons from the inside yields $(A^*, B, C, D^*, E, F, G, H, X, 5, 6, 8, 9)$, three variants, not applicable to the outside of 6 R 1 or 6 R 2. Making three pentagons with a common vertex from the inside yields $(A, B^*, C, D^*, E, F^*, G, H, R, S, T, U, V, W)$, two variants, not applicable to the outside of 6 R 1, 6 R 2, and 6 R 3.

Next we list all chain deductions from the first scheme in each cyclic group.

$$\begin{array}{ll}
 A \rightarrow (H, 1) \text{ and } \begin{pmatrix} C & E & X \\ R & W & 4 \end{pmatrix}. & X \rightarrow \begin{pmatrix} A & C & E \\ D & F & B \end{pmatrix}: \\
 A^* \rightarrow (H, 1) \text{ and } (U, 9) \text{ and } (T, 8). & 0 \rightarrow (1, 7) \text{ and } (2, 8) \text{ and } (3, 9). \\
 G \rightarrow (B, B^*) \text{ and } (D, D^*) \text{ and } (F, F^*). & 1 \rightarrow (A, A^*) \text{ and } (D, D^*) \text{ and } (0, 4). \\
 R \rightarrow (D^*, E^*) \text{ and } \begin{pmatrix} A & W & 4 \\ S & B & 5 \end{pmatrix}. & 4 \rightarrow (1, 7) \text{ and } \begin{pmatrix} R & A & W \\ U & D & T \end{pmatrix}: \\
 & 7 \rightarrow (B^*, C^*) \text{ and } (E^*, F^*) \text{ and } (0, 4).
 \end{array}$$

The ring pattern odd-even-odd-even-odd-even occurs for the first time in 6-rings and it involves more than two alternatives. Thus $X : 121212$ requires a selection of at least one scheme from each of six three-way choices (A, B, C) (B, C, D) (C, D, E) (D, E, F) (E, F, A) (F, A, B) . We indicate all six of these by the matrix $\begin{pmatrix} A & C & E \\ D & F & B \end{pmatrix}$. These requirements can be satisfied in any one of five sufficient ways (AD) (BE) (CF) (ACE) and (BDF) , namely by any row or column from the matrix.

The entire system of contingent options can be presented in a more compact manner. We may arrange the schemes which involve simple odd-even-odd-even ring patterns in groups of four,

AH	BG	CH	DG	EH	FG	10	20	30
1 A^*	2 B^*	3 C^*	1 D^*	2 E^*	3 F^*	47	58	69
B^*V	C^*W	D^*R	E^*S	F^*T	A^*U			
7 C^*	8 D^*	9 E^*	7 F^*	8 A^*	9 B^*			

in which each scheme implies another scheme at an adjacent corner. Those schemes which involve a threefold odd-even sequence around the ring we arrange in an array of 4 rows and columns corresponding to the design:

121212	3232	3212	1232	<i>X A C E</i>
3432	1412	1432	3412	<i>4 D U T</i>
3214	1234	1214	3234	<i>6 V F W</i>
1434	3414	3434	1414	<i>5 S R B</i>

Each element in this square array is to be thought of as belonging to one row and to one diagonal. To obtain the 2 by 3 option matrix for any scheme we write the elements in the same diagonal below the elements in the same row, retaining the column pairings. Or we may read the contingencies directly from this square array. Thus scheme *R* implies one of five sufficient alternatives: all the schemes (*5 S B*) in the same *R*-row; all the schemes (*4 A W*) in the same *R*-diagonal; or one representative from the *R*-row and one from the *R*-diagonal chosen from the same column (*4 5*) (*A S*) or (*W B*).

The next task is to find all solutions for the system of primary and contingent options. From any one solution we can, of course, produce others by cyclic permutation, by clock symmetry, and by interchanging inside and outside. We shall count all these variations as belonging to the same solution. We are now in a position to state the theorem: *The 6-ring options admit only these six solutions:*

Solution #1 (1 inside variation, 10 outside variations)

Inside *A B C D E F X 0 1 2 3*

Outside *A * B * C * D * E * F * G H 4 5 6 7 8 9* and two pairs (*R U*) (*S V*) (*T W*)
or more.

Solution #2 (3 variations, each side)

Inside *A * B C D * E F G H X 5 6 8 9*

Outside *A B * C * D E * F * R T U W 0 1 2 3 4 7*

Neither side *S V* .

Solution #3 (2 variations, each side)

Inside *A B * C D * E F * G H R S T U V W*

Outside *A * B C * D E * F X 0 1 2 3 7 8 9*

Neither side *4 5 6* .

Solution #4 (3 variations, each side)

Inside $AB*CD*EF*GHSVX456789$

Outside $A*BCD*EFTUW0123$.

Solution #5 (6 variations, each side)

Inside $AB*CD*EF*GHRTUW47$

Outside $A*BC*DE*FSVX01235689$.

Solution #6 (12 variations, each side)

Inside $ABC*DE*F*GHR SX4578$

Outside $A*B*CD*EFTUVW012369$.

Comments on these solutions: Solution #1 corresponds to a single hexagon inside; solution #2 corresponds to two pentagons inside, whose common edge in variation one is guarded by A and D ; solution #3 corresponds to three pentagons inside, with a common inside vertex, whose edges in variation one are guarded by A, C, E . That is, we know the inside structure of the first three solutions and it is a simple matter to verify that such structures are 4-colorable only for the schemes listed. We do not know the outside structure, and it is a matter of conjecture whether such structures, colorable as indicated, actually exist.

An attempt has been made to list the solutions in the order of their simplicity, but since the geometric structure of the last three solutions is unknown, the assigned order is merely for reference. Likewise the distinction between inside and outside is arbitrary. Solution #5 (with schemes 5 and 6 inadvertently omitted) was given by Birkhoff.³ He was disappointed in finding more than three solutions. But now that we know all the color solutions for the 6-ring, we can settle many questions without knowing their inside geometrical structure. In examining one side of a 6-ring we may consider in turn each of the 63 possible variations. Symmetry or other considerations may reduce the number of cases requiring separate study. The important contribution is that we have a tool for investigation—a partial “analytic geometry” for maps—which frees us from the necessity of drawing diagrams. We use geometric intuition to set up the options, but their simultaneous solution is a problem in permutations and combinations. Furthermore, our solutions are *complete* in the sense that each satisfies all the conditions proposed; and so no further results can be obtained from geometric devices which do not arrive at new conditions beyond those contained in our options. Our solutions are incomplete in that our options may not express all the conditions inherent in the geometric formulation of the problem.

Since the 6-ring theorem lists *all* solutions, we may set down a number of simple corollaries which summarize their characteristics. Observe that A and A^* are always on opposite sides; and that (G, H) are always opposite $(0, 1, 2, 3)$. Also notice the absentees 4, 5, 6 in #3, and SV in #2, optionally absent in #1. The only solution clock-asymmetric is #6. The only solution with three-phase symmetry is #3.

Inside Structure for 6-rings—Preliminary Considerations.

Let $6Rn$ indicate that a ring of six regions surrounds an inside group of n regions. It has already been mentioned that the first three solutions of the 6-ring correspond to $6R1$, $6R2$ and $6R3$ respectively.

If both $6Rn'$ and $6Rn''$ have the same 6-ring solution, then $n' = n''$. For if n' is not greater than n'' then the inside with n' regions could be matched with the outside with $m_0 - n'' - 6$ regions forming a non-colorable map with $m = m_0 - (n'' - n')$ regions. But m_0 is the minimal value of m ; therefore $n' = n''$. Except for the ambiguous $6R1$ outside, if a geometric structure can be found fitting one variation, it could be rotated and reflected to fit all the other variations. It is conceivable that more than one geometric structure (different arrangements of the same number of regions) might fit the same solution, or that no such structure was geometrically possible. Again, such structures might be possible, yet not essential, features of every minimal map.

Knowing the color solutions of the six possible 6-rings, it is natural to seek geometric structures meeting these requirements. But the trial and error method of examining one geometric candidate after another seems to be unfruitful. Birkhoff³ cautiously confessed: "In all the cases which I have considered, the ring R of six regions containing more than three [inside] regions . . . is reducible." (The terminology *reducible* means that the problem of coloring such a map reduces to the problem of coloring another map with fewer regions. In minimal maps reducible configurations simply do not occur.) We propose an approach to 6-ring structure by means of a cumulative series of questions. Excluding from consideration the $6R1$ inside, $6R2$ inside, and $6R3$ inside, whose structures are already completely known, our questions will apply to each side of all six rings. Each such side must be colorable according to one of the variations listed as possible in the foregoing 6-ring theorem. Since the color schemes of each variation are known, we can decide whether its colorability is compatible with the conditions in question. When a like verdict is obtained for each of the variations, we have obtained a theorem valid for all 6-rings. If the verdict should depend upon

the variation, then we have a clue to the characteristics which distinguish one variation from another. Each successive answer will suggest the formulation of new questions. This is our program. The questions discussed in this paper are by no means exhaustive, but they will serve to illustrate the fruitfulness of this approach.

Can any inside region Y make multiple contact with the 6-ring? To avoid taboo 2, 3, 4-rings contact must be at opposite sides, say A and D . Then $ABCDY$ and $AYDEF$ are 5-rings, so that their insides X and Z are single regions. But then $AXDZ$ forms a 'taboo 4-ring around Y . Hence *multiple contact is excluded*, and the peripheral cycle of inside regions making contact with the ring is a *distinct set*.

Can any inside region contact as many as three ring regions? Since multiple contact has been excluded, the contact must be consecutive. If D' contacts $ABCDE$ then $AD'EF$ threatens to form a 4-ring unless D' is the only inside region, hence 6 R 1. If D' contacts only $ABCD$ then $AD'DEF$ forms a 5-ring around D'' , and $D'D''$ are the two inside pentagons for 6 R 2. If D' contacts only CDE then both $ABCDEF$ and $ABCD'EF$ would be 6-rings with DD' forming a *double link*.

Designating the respective insides by I and J we observe that I contains one more region D' than J . But the colorability of I and J are interdependent. Postulating scheme C for one side implies T for the other side, and conversely. The same reciprocity holds for E and U , A^* and H , B^* and 2, C^* and 8, E^* and 9, F^* and 3, S and 6, V and 5, 0 and 7. Of the three schemes B, F, G postulating one for I implies an option of the other two for J , while postulating one for J implies both the other two for I . (Consequently postulating one of the three schemes for I implies a second for I , and the third for J .) The same implications hold for the triads $A, 1, 4$ and D^*, R, W . Finally postulating X for one side implies D for the other, but $I(D) \rightarrow J(D, X)$ while $J(D) \rightarrow I(D, X)$, relations which could be obtained formally by treating D, D, X as a fourth triad.

In view of these interdependent relations between I and J we may answer the double-link question by systematically trying all 6-ring solutions seeking an IJ fit.

$J(0, 1, 2, 3) \rightarrow J(0. 1. 2. 3. \text{ not-}G. \text{ not-}H) \rightarrow I(AB^*F^*4\gamma) \rightarrow J(B^*F^*)$ since B^* is equivalent to not- B , F^* to not- F . Thus we can ignore any postulate $J(0, 1, 2, 3)$ not coupled with $J(B^*F^*)$. The alternate postulate $J(GH) \rightarrow I(BFA^*)$ whence J is not-4. This excludes many J cases, in particular both sides of 6 R 1 and both sides of #4. Since C, C^* never occur simul-

taneously, we avoid putting T and 8 on the same side, J or I ; likewise $E.E^*$ indicates U and 9 are incompatible; while $D.D^*$ indicates X or D should not be placed on J with either R or W . Only six cases for J pass these preliminary hurdles, and we examine these individually.

First case: J is #2 inside, first variation, with $A^*BCD^*EFGHX5689$. Then I is $A^*BC^*DE^*FGHRSUVW$, which is #3 inside, second variation.

Second case: J is #3 inside, first variation, with $AB^*CD^*EF^*GHRSTUVW$. Then I is $A^*BCD^*EFRTUW1456$ which does not fit any 6-ring solution.

Third case: J is #3 outside, second variation, with $AB^*CD^*EF^*X0123-789$. Then I is $AB^*C^*DE^*F^*TU0147$ which fits #2 outside, first variation. This is the first case as seen from the other side.

Fourth case: J is #5 inside, third variation, with $AB^*CD^*EF^*GHTV-WS69$. Then I is $A^*BCD^*E^*FRSTUW123456$, which does not fit any 6-ring solution.

Fifth case: J is #5 inside, fifth variation, with $AB^*CD^*EF^*GHSUV58$. Then I is $A^*BC^*D^*EFRTUW123456$ which does not fit.

Sixth case: J is #5 outside, fourth variation, $AB^*CD^*EF^*SVX01235689$. Then I is $AB^*C^*DE^*F^*RSTUVW1234567$, which does not fit.

Summary: In any 6 R 3 inside, each of the enclosed pentagons produces a double link, forming three associated 6 R 2 structures. Otherwise, 6-rings do not possess double links; no region contacts more than two ring regions. The regions in the outermost I shell belong to two types: *corner regions* which contact two ring regions, guarding their common edge; and *lateral regions* which contact only one ring region. Incidentally, since this shell certainly contains six corners, 6 R 4 and 6 R 5 structures are excluded.

If there are p regions in this peripheral I shell (p -shell: composed of *corner* regions and *lateral* regions) and h additional regions on the inside, then the average number of sides for inside regions exceeds 5 by $\frac{h}{p+h}$, so that if they were only pentagons and hexagons there would be p pentagons and h hexagons.

Can non-consecutive regions in the p -shell be adjacent? Having shown that the regions in the outermost shell are distinct, we show that they constitute a ring! The only contact between non-consecutive regions in the shell not forming taboo 3-rings, 4-rings, or 5-rings would be between lateral

regions at opposite sides of the hexagonal contour. Label the corner regions R at AB , S at BC , T at CD , U at DE , V at EF , W at FA . Let A' be in lateral contact with A , and D' with D . Postulating A' adjacent to D' forms a *system of three 6-rings*: $ABCDD'A'$ with RST inside, $AA'D'DEF$ with UVW inside, and $AFEDCB$ with $WA'R$ outside. Designate the respective insides by I, J, K . *Their colorability is interdependent.* Thus $I(A)$ and $J(C)$ together imply $K(\text{not-}E)$ else the entire map would be 4-colorable. In the ordered triad ACE , postulating the first scheme for I and the second scheme for J implies the exclusion of the third scheme from K . Symmetry in the definition of I, J, K allows the corresponding inferences from the triads CEA and EAC , (but not from ECA). 001 forms another such triad, and since the schemes 0 and 1 are always opposite H , we conclude that at least one of the sides I, J, K is colorable H . We next try to assign color schemes to I, J, K consistent with the six known solutions but avoiding respective assignments which involve taboo triads such as $I(A).J(C).K(E)$. The method is simple but involves a consideration of many cases, and we omit the details here. The conclusion from the systematic trials is that no fit is possible: the postulate A' adjacent to D' is absurd. Therefore *the p -shell constitutes a p -ring.*

In an entirely similar manner we may dispose of the question: *Can opposite corner regions W and T have a common neighbor X ?* Another system of 6-rings is formed: $DEFWXT$, $TXWABC$, and $CBAFED$. Another system of scheme triads is formed such as OAD and again we may try for a fit. No fit can be made and the postulated structure is impossible.

Since the six or more regions in the first shell form a ring each must be adjacent to some region in the second shell. Since the same region X cannot serve for both W and T there must be at least two regions in the second. (For $p=6$ we may apply our 6-ring conclusions to the p -ring. The trials $pR1$, $pR2$, $pR3$ collapse, whence there would be a p' -ring within the p -ring.) That makes at least nine regions inside the original 6-ring, so we have *excluded all 6 R 6, 6 R 7 and 6 R 8 structures.*

Can the p -shell be another 6-ring, contiguous with the original 6-ring and consisting of corner regions only? Let the inside I of $ABCDEF$ include R and let the inside J of $RSTUVW$ exclude A . Then the color schemes of I and J are interdependent. Thus $J(B^*) \rightarrow I(A^*B^*C^*D^*0189)$, but since no 6-ring solution contains 0 together with four "consecutive star" schemes, the postulate $J(B^*)$ is absurd, and by symmetry *all* star schemes are excluded from J . But then J would have to be the 6 R 1 inside, colorable by scheme 0. But $J(0)$ leads to the impossible $I(H.0)$, so that *contiguous 6-rings do not occur.*

Can the p -shell be a 7-ring? We suppose a region A' making lateral contact with A , plus the six corner regions. Then $A'RSTUVW$ is a 7-ring contiguous with the 6-ring. Considering each of the 63 cases, the color schemes not possible for that side of the 6-ring including A' induce restraints on the coloring of the 7-ring. But too many restraints become inconsistent with the requirements of certain primary options for the 7-ring. In this manner we may dispose of each case. Their verdict is unanimous, so that *a 7-ring contiguous with a 6-ring does not occur.*

In like manner minimal maps reject the occurrence of an 8-ring contiguous with a 6-ring. The 8-ring possesses two lateral regions in addition to the six corner regions. We distinguish four structures: (1) A' and A'' sharing a vertex with A , (2) A' and B' in lateral contact with A and B , respectively, (3) A' and C' in lateral contact with A and C , respectively, and (4) A' and D' in lateral contact with A and D , respectively. An exhaustive examination of each case shows that no color fit is possible, and the proposed structure is excluded.

We conclude our current study of the inside structure of 6-rings with the question: *Can alternate corner regions S and U have a common neighbor X in the second shell?* If so, then $UXSCD$ forms a 5-ring around the pentagon T . The colorability of the 7-ring $ABSXUEF$ and of the 6-ring $ABCDEF$ are interdependent. Let I be that side of the 6-ring which excludes X , and let J be that side of the 7-ring which excludes T . To each scheme for which I can be colored there are certain schemes excluded from J , lest the entire map be 4-colorable, contrary to our basic conjecture. Usually each I scheme excludes four J schemes, but $B^*, E^*, 0, 2, 7, 9$ each exclude six J schemes, while $I(X)$ excludes only two J -schemes. An examination of all 63 cases for I shows that only one case is compatible with the supposed structure, solution #1 inside, that in which I is a single hexagon. Then C and D are pentagons, the primary components of an edge guarded by the pentagon T and the hexagon I .

Franklin⁵ has shown that three pentagons and one hexagon cannot form an edge with the hexagon in primary position. But when the hexagon is in guard position, the $7R4$ structure is *irreducible* since its outside may be assigned color schemes compatible with all known primary and contingent options.

The reader is reminded that these cumulative results concerning the

⁵ P. Franklin, *American Journal of Mathematics*, vol. 44 (1922), pp. 225-236.

"inside structure" of 6-rings apply to *both* sides of solutions #4, #5, #6 and—with specified exceptions—to the *outside* of solutions #1, #2, #3.

Neighboring the original six-ring R are two more rings each containing at least nine regions. The p regions of each associated ring consist of six *corner* regions (guarding the edges between consecutive R regions) plus three or more *lateral* regions. Inside the p -ring a corner n -gon contacts $n - 4$ regions, while a lateral n -gon contacts $n - 3$ regions. Opposite corner regions have no common neighbor, and alternate corner regions have no common neighbor inside the p -ring. Defining the power of an n -gon as $n - 5$, the combined power of the $p + h$ regions inside the 6-ring is precisely h . By resorting to geometrical trials it can be shown that the power h cannot be concentrated in five or fewer regions, nor yet in six hexagons. This includes the (weaker) corollary that h is greater than six. Combined with the knowledge that p is at least nine, we see that $p + h$ exceeds fifteen. Thus $6Rn$ structures with $3 < n < n_0 = 16$ do not exist in minimal maps. The way is open to increase the value of p , h and n_0 . Birkhoff³ conjectured that six-rings not surrounding one, two or three regions might be a characteristic feature of all minimal maps. If one such ring is present in one minimal map of m_0 regions then m_0 is not less than $2n_0 + 6$.

Proof of the Main Theorem.

The superscript notation relative to scheme S provides for five classes:

S^1 inside colorable S^3 outside not-colorable $\therefore (S^1, S^5)$

S^2 outside colorable S^4 inside not-colorable $\therefore (S^2, S^5)$

S^5 neither side colorable

$$(RST)^2 = R^2.S^2.T^2$$

The problem is solved when each of the 31 schemes has been assigned one of the superscripts 1, 2, or 5 consistent with the primary and contingent options.

For convenient reference we tabulate herewith all the *primary* options:

- | | |
|-----------------------|-----------------------------|
| (1) (A, C, E, H, X) | (21) $(A, D^*, W, R, 1, 4)$ |
| (2) (B, D, F, G, X) | (22) $(B, E^*, R, S, 2, 5)$ |
| | (23) $(C, F^*, S, T, 3, 6)$ |
| (3) $(A, D, X, 1, 4)$ | (24) $(D, A^*, T, U, 1, 4)$ |
| (4) $(B, E, X, 2, 5)$ | (25) $(E, B^*, U, V, 2, 5)$ |
| (5) $(C, F, X, 3, 6)$ | (26) $(F, C^*, V, W, 3, 6)$ |

- | | |
|-------------------------------------|---|
| (6) $(A^*, D^*, 0, 1, 8, 9)$ | (27) (A^*, C, E, H, T, U) |
| (7) $(B^*, E^*, 0, 2, 7, 9)$ | (28) (B^*, D, F, G, U, V) |
| (8) $(C^*, F^*, 0, 3, 7, 8)$ | (29) (C^*, E, A, H, V, W) |
| | (30) (D^*, F, B, G, W, R) |
| (9) $(A^*, D^*, R, U, 1, 4, 5, 8)$ | (31) (E^*, A, C, H, R, S) |
| (10) $(B^*, E^*, S, V, 2, 5, 6, 9)$ | (32) (F^*, B, D, G, S, T) |
| (11) $(C^*, F^*, T, W, 3, 4, 6, 7)$ | |
| | |
| (12) $(A^*, D^*, T, W, 1, 4, 6, 9)$ | (33) $(B, E, C^*, E^*, W, R, 0, 2, 3, 7)$ |
| (13) $(B^*, E^*, R, U, 2, 4, 5, 7)$ | (34) $(C, A, D^*, F^*, R, S, 0, 1, 3, 8)$ |
| (14) $(C^*, F^*, S, V, 3, 5, 6, 8)$ | (35) $(D, B, E^*, A^*, S, T, 0, 1, 2, 9)$ |
| | (36) $(E, C, F^*, B^*, T, U, 0, 2, 3, 7)$ |
| (15) $(A, C^*, E^*, H, W, R, 4, 7)$ | (37) $(F, D, A^*, C^*, U, V, 0, 1, 3, 8)$ |
| (16) $(B, D^*, F^*, G, R, S, 5, 8)$ | (38) $(A, E, B^*, D^*, V, W, 0, 1, 2, 9)$ |
| (17) $(C, E^*, A^*, H, S, T, 6, 9)$ | |
| (18) $(D, B^*, F^*, G, T, U, 4, 7)$ | |
| (19) $(E, A^*, C^*, H, U, V, 5, 8)$ | |
| (20) $(F, B^*, D^*, G, V, W, 6, 9)$ | |

The contingent options have been arranged in compact "color matrices"

$XACE$	AH	BG	CH	DG	EH	FG	10	20	30
4DUT	1A*	2B*	3C*	1D*	2E*	3F*	47	58	69
6VFW	RE*	SF*	TA*	UB*	VC*	WD*			
5SRB	D*9	E*7	F*8	A*9	B*7	C*8			

which will be indicated by enclosing the scheme found in the upper left corner in quotes, thus "X," "A," etc. We are now in a position to seek systematically all the solutions of this combinatorial problem.

Case 1. Excluding all "star" schemes from one side let $(A^*B^*C^*D^*E^*F^*)^4$. Then "R," "S," . . . "W" imply $(RSTUVW789)^4$. Primary (6) yields the option $(0, 1)^1$ but $(0)^1(7)^4$ yields $(1)^1$ by "1"; therefore $(0, 1)^1$ gives $(1)^1$. Similarly, primaries (7) and (8) give $(2)^1$ and $(3)^1$. We already have assigned three schemes to the inside, namely $(123)^1$. Now $(123)^1(\text{stars})^4$ via "A," "B," etc. gives $(ABCDEF)^1$, and $(A)^1(WR)^4$ via "X" gives $(X)^1$. Primary (1) with $(ACEX)^1$ gives $(H)^2$, similarly primary (2) gives $(G)^2$. Now $(H)^2(A)^3$ via "A" gives $(A^*)^2$; by symmetry this applies to all the stars, thus $(A^*B^*C^*D^*E^*F^*)^2$. Primary (3) with $(ADX1)^1$ gives $(4)^2$, similarly (4) and (5) give $(5)^2$ and $(6)^2$. Now $(4)^2(1)^1$ via "1" gives

$(7)^2(0)^1$; similarly $(5)^2$ and $(6)^2$ give $(8)^2$ and $(9)^2$. We already have assigned $ABCDEFX0123$ to the inside and $A^*B^*C^*D^*E^*F^*GH456789$ to the outside. Finally " X " with $(4)^2$ provides the outside option $(RU, TW)^2$ while $(5)^2$ and $(6)^2$ provide the outside options $(RU, SV)^2$ and $(SV, TW)^2$. These compel two pairs outside, for example RU and TW . All requirements are fulfilled so S and V may be assigned in any of four ways $(SV)^2$ or $(SV)^5$ or S^2V^5 or S^5V^2 . This is solution #1, the only possibility under Case 1.

Case 2. Excluding five star schemes from one side, let $(A^*B^*C^*D^*E^*)^4$. Excluding Case 1 [already considered] we have $(F^*)^1$. From " R ," " U ," " V ," and " W " we have $(789RUVW)^4$. Now $(F^*)^1(78)^4$ via " S " and " T " yields $(S)^1$ and $(T)^1$ respectively. Primary (6) provides the option $(0, 1)^1$ which combines with $(7)^4$ via " 1 " to give $(1)^1$. Similarly primary (7) gives $(2)^1$. Now $(1)^1(A^*D^*)^4$ gives $(A)^1$ and $(D)^1$ via " A " and " D " respectively. Likewise $(2)^1(B^*E^*)^4$ gives $(B)^1$ and $(E)^1$ via " B " and " E ". Primary (32) gives $(G)^2$, which with $(BDF^*)^1$ via " B " " D " and " F " gives $(B^*D^*F)^2$. Primary (20) gives $(6)^1$, and $(6)^1(9)^4$ via " 3 " gives $(3)^1$; and $(3)^1(C^*)^4$ via " C " gives $(C)^1$. We now have $(CF^*ST36)^1$ contradicting primary (23). Thus Case 2 falls.

Case 3. Excluding four consecutive star schemes from one side, let $(A^*B^*C^*D^*)^4$. Excluding Case 2 [already considered] we have $(E^*F^*)^1$. From " U ," " V " and " W " we have $(UVW789)^4$ and now $(E^*F^*)^1$ via " R ," " S " and " T " gives $(RST)^1(7)^5$. Primary (6) provides the option $(0, 1)^1$ which combines with $(7)^4$ via " 1 " to give $(1)^1$. Now $(1)^1(A^*D^*)^4$ via " A " and " D " gives $(AD)^1$. At this point $(BC)^2$ would be absurd. For $(BC)^2(B^*C^*)^4$ via " B " and " C " would imply $(GH23)^4$ which is incompatible with $(E^*)^1$ by virtue of " E ". Accordingly $(B, C)^3$ and from the symmetry we may choose $(C)^3$.

Then primary (31) gives $(H)^2$, and then $(H)^2(ACE^*)^3$ via " A ," " C " and " E " gives $(A^*C^*E)^2$. Now $(E)^2(AC)^3$ via " X " gives $(5)^2$.

Further $(C^*)^2(7)^3$ via " V " gives $(V)^2$. But then $(EA^*C^*HUV58)^4$ is incompatible with primary (19). Thus Case 3 falls.

Case 4. Let $(A^*B^*C^*)^4(E^*)^3$. Avoiding case 3 already considered we have first $(D^*F^*)^1$ and then $(A^*C^*)^2$. Thus case 4 has the symmetrical form $(D^*F^*)^1(A^*C^*)^2$. $(E^*)^3(B^*)^4$. Now $(A^*B^*C^*)^4$ via " U " and " V " gives $(UV79)^4$, while symmetry gives $(RS79)^3$, therefore $(79)^5$. Again $(D^*F^*)^1(79)^5$ via " R " and " S " gives $(RS)^1$, while symmetry gives $(UV)^2$. At this point $(H)^1$ is absurd for $(H)^1(A^*C^*)^2$ via " A " and " C " would

give $(AC)^1$, but the resulting $(E^*ACHRS)^3$ is incompatible with primary (31). This absurdity of $(H)^1$ establishes $(H)^4$, and symmetrically $(G)^3$.

Since scheme 2 cannot be on both sides, either $(2)^3$ or $(2)^4$. We remove symmetry by choosing $(2)^3$. With $(G)^3$ via "B" this gives $(BB^*)^3$. Primary (7) yields $(0)^2$, while $(0)^2(729)^3$ via "1," "2," and "3" gives $(183)^2$. Now $(13)^2(D^*F^*)^1$ via "D" and "F" compels $(DF)^2$, but the resulting $(FDA^*C^*UV0138)^2$ is incompatible with primary (37). Thus case 4 falls.

Case 5. Excluding three consecutive stars from one side, let $(A^*B^*C^*)^4$. Fallen case 3 requires $(D^*F^*)^1$, and fallen case 4 requires $(E^*)^2$. From $(A^*B^*C^*)^4$ via "U" and "V" we have $(UV79)^4$ which in turn combines with $(D^*F^*)^1(E^*)^2$ via "R" and "S" giving $(RS)^1(79)^2$.

If $(0)^4$ then primary (7) would require $(2)^1$. But this would give $(5)^1$ and $(E)^1$ via "2" and "E" respectively; and $(5E)^1(UV)^1$ via "X" would give $(BX)^1$.

But $(BEX25)^1$ is incompatible with primary (4). The absurdity of $(0)^4$ establishes $(0)^1$, and $(0)^1(79)^2$ via "1" and "3" establishes $(13)^1$ which in turn combines with $(A^*C^*)^4$ via "A" and "C" to give $(AC)^1$. Now $(ACD^*F^*RS013)^1$ and primary (34) yields $(8)^2$, and $(8)^2(0)^1$ via "2" gives $(5)^2(2)^1$. Again $(2)^1(E^*)^4$ via "E" gives $(E)^1$. But $(5)^2$ via "X" is incompatible with $(ERS)^1$. Thus case 5 falls. In any additional solutions each side must contain at least one of any three consecutive stars.

Case 6. Let $(A^*C^*E^*789)^2$.

Fallen case 5 compels $(B^*D^*F^*)^1$ which combines with $(789)^2$ via "R" . . . "W" giving $(RSTUVW)^1$.

At this point $(G)^2$ is absurd. For $(G)^2(B^*D^*F^*)^1$ via "B," "D" and "F" gives $(BDF)^2$, and $(B)^2(RS)^1$ via "X" gives $(X)^2$. But $(BDFGX)^2$ is incompatible with primary (2). The absurdity of $(G)^2$ establishes $(G)^3$.

Using $(B^*D^*F^*RSTUVW)^1(G)^3$ primary options (28), (30) and (32) provide the three outside options $(D, F)^2$ and $(F, B)^2$ and $(B, D)^2$. Thus the outside contains at least two of the three schemes B, D, F and we may exploit the symmetry by choosing $(BD)^2$. Then $(BD)^2(G)^3(B^*D^*)^1$ via "B" and "D" gives $(12)^2(G)^1$. Also $(1278)^2$ via "1" and "2" gives $(045)^4$. Now $(R)^1(B5)^4$ via "X" gives $(A)^1$, and $(A)^1(1)^2$ via "A" gives $(II)^1$. In continuation $(H)^1(C^*E^*)^2$ via "C" and "E" gives $(CE)^1(3)^2$, and $(3)^2(F^*)^1$ via "F" gives $(F)^2$, restoring the symmetry, whence also $(6)^4$. Finally $(ATU)^1$ and $(ERS)^1$ and $(CVW)^1$ via "X" require $(456)^3$

which combines with $(456)^4$ to give $(456)^5$. The assignment $(0X)^2$ is completed by $(1)^2(4)^5$ via "1," and by $(B)^2(RS)^1$ via "X," respectively.

The result

$$(ACEB^*D^*F^*GHRSTUVW)^1(BDFA^*C^*E^*X0123789)^2(456)^5$$

satisfies all requirements. This is solution #3.

Case 7. Excluding alternate stars from one side, let $(A^*C^*E^*)^4$. Fallen case 5 compels $(B^*D^*F^*)^1$ and $(A^*C^*E^*)^2$. From "R" . . . "W" the contingencies (7, SV) and (8, TW) and (9, RU) apply to both sides. Assigning 7·8·9 to the same side is case 6, so we seek new solutions by assigning two of these schemes to one side, the third scheme to the other side. Symmetry makes $(78)^2(9)^1$ typical. We have therefore $(A^*C^*E^*78RU)^2(B^*D^*F^*9STVW)^1$. Then $(R)^2(SW)^1$ via "X" implies $(45)^2$.

An inspection of "X" reveals that $(X)^3$ is absurd. For $(STX)^3$ implies $(C)^3$; $(VWX)^3$ implies $(F)^3$; $(CVW)^3$ implies $(6)^3$; and $(FF^*)^3$ via "F" implies $(3)^3$. But $(CFX36)^3$ is incompatible with primary (5): Accordingly $(X)^2$.

It is readily shown that both $(6)^2$ and $(G)^2$ are absurd. On the one hand $(6)^2(SVW)^1$ via "X" gives $(CF)^2$, and $(6)^2(9)^1$ via "3" gives $(3)^2$. But $(CFX36)^2$ is incompatible with primary (5). Therefore $(6)^3$. On the other hand $(G)^2(B^*D^*F^*)^1$ via "B," "D" and "F" gives $(BDF)^2$, but $(BDFGX)^2$ is incompatible with primary (2). Therefore $(G)^3$.

Now however primary (20) requires $(F)^2$, which combines with $(6VW)^3$ via "X" to give $(BD)^2$, and in continuation $(BDF)^2(G)^3$ via "B," "D" and "F" gives $(123)^2(G)^1$. Again $(3)^2(6)^3(9)^1$ via "3" gives $(0)^2(6)^1$. If $(H)^4$ then $(123)^2$ via "A," "C" and "E" would compel also $(ACE)^4$, but the resulting $(ACEHX)^4$ is incompatible with primary (1). Therefore $(H)^4$ is absurd, and instead we have $(H)^1$. Now $(H)^1(A^*C^*E^*)^2$ via "A," "C" and "E" gives $(ACE)^1$. The complete assignment of schemes

$$(ACEB^*D^*F^*GHSTVW69)^1(BDFA^*C^*E^*RUX01234578)^2$$

satisfies all requirements. It is solution #5.

The seven cases already investigated have served to indicate three solutions (listed as #1, #3 and #5). By virtue of case 5 and case 7 any further solutions must be obtained by assigning to each side at least one scheme from each consecutive group of three stars and at least one from each group of three alternate stars.

Case 8. Let $(B^*C^*E^*F^*)^1(A^*D^*89)^2$. The hypothesis of this case

combines with the contingencies " R " . . . " W " to give $(RTUW)^1(SV\gamma)^3$. We readily see that $(0)^2$ is absurd, for $(0)^2(\gamma)^3$ via " 1 " implies $(1)^2$, but $(A^*D^*0189)^2$ is incompatible with primary (6). Therefore $(0)^3$, which combines with $(89)^2$ via " 2 " and " 3 " to give $(56)^2$. Then $(5)^2(SU)^3$ and $(6)^2(TV)^3$ via " X " imply $(BE)^2$ and $(CF)^2$ respectively. Also $(B)^2(RS)^3$ via " X " implies $(X)^2$. Again $(BEX5)^2$ and primary (4) yield $(2)^1$, while $(CFX6)^2$ and primary (5) yield $(3)^1$. Now $(BC)^2(23)^1$ via " B " and " C " yields $(GH)^2$, and $(2)^1(5)^2$ via " 2 " yields $(0)^1$. An inspection of " X " shows that $(BC5)^2$ infers S^4 , and that $(EF6)$ infers V^4 . These results combine with $(SV)^3$ to give $(SV)^5$, while $(E^*)^1(S)^5$ via " S " gives $(7)^1$.

Finally $(CEHX)^2$ and primary (1) compel $(A)^1$, while $(BFGX)^2$ and primary (2) compel $(D)^1$, so that $(A)^1(H)^2$ via " A " gives $(1)^1$, and $(D)^1(BF)^2$ via " X " yields $(4)^1$. The completed assignment

$$(ADB^*C^*E^*F^*RTUW012347)^1(BCEFA^*D^*GHX5689)^2(SV)^5$$

satisfies all requirements. This is solution #2.

Case 9. Let $(B^*C^*E^*F^*89)^1(A^*D^*)^2$. The hypothesis of this case combines with the contingencies " R " . . . " W " to give $(RTUW)^2(SV\gamma)^3$. Primaries (7) and (8) provide the outside options $(0, 2)^2$ and $(0, 3)^2$ respectively. Accordingly $(23)^2$ is compulsory, since it follows both from $(0)^3$ via these options, and from $(0)^2$ with $(89)^1$ via " 2 " and " 3 ." Then $(2)^2(B^*E^*)^1$ via " B " and " E " gives $(BE)^2(GH)^1$, and $(3)^2(C^*F^*)^1$ via " C " and " F " gives $(CF)^2$. In continuation $(H)^1(A^*)^2$ via " A " gives $(A)^1(1)^2$, while $(G)^1(D^*)^2$ via " D " gives $(D)^1$. Via " X " from $(A)^1(RW)^2$ there follows $(X)^1$, and from $(D)^1(BF)^2$ follows $(4)^1$. Now $(1)^2(4)^1$ via " 1 " gives $(0)^2(\gamma)^1$, and $(89)^1(0)^2$ via " 2 " and " 3 " gives $(56)^1$. Finally $(5)^1(BU)^2$ via " X " gives $(SV)^1$. The result

$$(ADB^*C^*E^*F^*GHSVX456789)^1(BCEFA^*D^*RTUW0123)^2$$

satisfies all requirements. This is solution #4.

Case 10. Excluding the special cases comprised under case 8 and case 9 let us consider $(B^*C^*E^*F^*)^1(A^*D^*)^2$. The contingencies " R " . . . " W " provide the restraint $(SV\gamma)^3$, and the options (8, TW) and (9, RU) applicable to both sides. Accordingly 8 and 9 must each be assigned to one side or the other. Putting 8·9 both outside is case 8; and putting 8·9 both inside is case 9. There remains only the possibility of putting one inside, the other

outside. As the two choices are symmetrically equivalent, let us consider $(8)^1(9)^2$ whence $(TW)^2(RU)^1$. Under case 10 we therefore examine

$$(B^*C^*E^*F^*RU8)^1(A^*D^*TW9)^2(SV7)^3.$$

Primary (8) reduces to the outside option $(0, 3)^2$. The trial $(0)^1$ is contrary to this option since $(0)^1(9)^2$ via "3" gives $(3)^1$. Therefore $(0)^4$, and $(8)^1(0)^4$ via "2" gives $(5)^1$. The trial E^3 is likewise absurd. For $(EE^*)^3$ via "E" gives $(2)^3$, but $(EB^*UV25)^3$ is incompatible with primary (25). The absurdity of $(E)^3$ establishes $(E)^2$, which combines with $(UV5)^3$ via "X" to give $(ACX)^2$. Now $(AA^*)^2$ via "A" compels $(H)^4$, but $(ACEHX)^4$ is incompatible with primary (1). Thus case 10 falls.

Case 11. Putting three stars on each side, [but not alternately as in case 7] let $(A^*B^*D^*)^2(C^*E^*F^*)^1$.

From "T" and "W" we obtain the option $(8, TW)$ applicable to both sides, and we may remove the symmetry by choosing $(8)^1(TW)^2$. From $(E^*F^*)^1$ in "S" follows $(S7)^3$, and from $(A^*B^*)^2$ in "U" follows $(U9)^4$. But $(B^*)^2(7)^3(C^*)^1$ in "V" gives $(V)^2(7)^1$, and $(E^*)^1(9)^4(D^*)^2$ in "R" gives $(R)^1(9)^2$.

Try $(0)^3$. Primary (8) compels $(3)^2$. Then $(3)^2(C^*F^*)^1$ via "C" and "F" gives $(CF)^2(GH)^1$, and $(GH)^1(A^*B^*)^2$ via "A" and "B" gives $(AB)^1(12)^2$. Now $(1)^2(0)^3$ via "1" gives $(4)^2$, but $(A)^1(CW4)^2$ is incompatible in "X." Then the trial $(0)^3$ fails, and instead we have $(0)^2$.

Now $(0)^2(78)^1$ via "1" and "2" gives $(12)^2(45)^1$, and $(2)^2(E^*)^1$ via "E" gives $(E)^2(H)^1$, whence $(H)^1(A^*)^2$ via "A" gives $(A)^1$. At this point primary (31) compels $(C)^2$, and $(C)^2(H)^1$ via "C" gives $(3)^2$. Then $(3)^2(F^*)^1$ via "F" gives $(F)^2(G)^1$, and $(G)^1(B^*D^*)^2$ via "B" and "D" gives $(BD)^1$.

Examining "X," we find $(W)^2(AR)^1$ gives $(6)^2$, while $(A)^1(CW)^2$ gives $(X)^1$, whence $(E)^2(XA)^1$ gives $(U)^2$, and $(5)^1(EU)^2$ gives $(S)^1$. The result $(ABDC^*E^*F^*GHR SX4578)^1(CEFA^*B^*D^*TUVW012369)^2$ satisfies all requirements. It is solution #6.

Case 12. Avoiding three stars on each side there remains but one other pattern to consider. Let $(B^*E^*)^1(A^*D^*)^2(C^*)^3(F^*)^5$. For if one side has more than three stars, the other side [except in case 1] must have diagonal stars, and we have the pattern $(A^*D^*)^2(B^*C^*E^*F^*)^1$ the consideration of which was concluded in case 10. If neither side has more and one side has less than three stars we have the pattern $(A^*D^*)^2(B^*C^*E^*)^3(F^*)^5$ in which the selection $(F^*)^5$ merely removes the symmetry. But F^*, A^*, B^* and

D^*, E^*, F^* are consecutive stars, so that to avoid duplicating case 5 we must specify $(B^*E^*)^1$ as has been done.

From $(B^*C^*E^*F^*)^3$ via "S" and "V" we have $(SV\gamma)^3$. From $(A^*F^*)^4$ via "T" we have $(T8)^4$. From $(B^*E^*)^1(A^*D^*)^2$ via "R" and "U" we have the option $(9, RU)$ applicable to both sides.

We next show that $(2)^4$ is absurd. For $(2 \cdot 8)^4$ via "2" implies $(05)^4$ while $(B^*)^1(2)^4$ via "B" implies $(G)^1$, and $(G)^1(D^*F^*)^4$ via "D" and "F" implies $(DF)^1$. Now primary (28) compels $(U)^2$, and using the option $(9, RU)$ also $(R)^2$. Again, primary (2) reduces to $(B, X)^2$, but since $(B)^2(FS)^3$ via "X" implies $(X)^2$, therefore $(B, X)^2$ implies $(X)^2$. Then $(X)^2(DF)^1$ via "X" gives $(E)^2$. Meanwhile $(E^*)^1(2)^4$ via "E" gives $(H)^1$, and $(H)^1(A^*)^2$ via "A" gives $(A)^1$. However $(A)^1$ is incompatible with $(ERX)^2$ in "X." This absurdity establishes $(2)^1$.

Next examine $(9)^1(RU)^2$. Primary (7) yields $(0)^2$, and $(2)^1(0)^2$ via "2" gives $(5)^1(8)^2$, while $(9)^1(0)^2$ via "3" gives $(6)^1(3)^2$; and $(0)^2(7)^3$ via "1" gives $(1)^2$. Now $(3)^2(C^*)^3(F^*)^5$ via "C" and "F" gives $(CF)^2(G)^4$, and $(G)^4(1)^2$ via "D" gives $(D)^4$. With $(25)^1$ primary (4) reduces to the outside option $(B, E, X)^2$. Since $(X)^1(CDF)^4$ via "X" would give $(BE)^1$ voiding this option, therefore $(X)^4$. Then $(DRX)^4$ via "X" compels $(B)^4$. But $(BDFGX)^4$ is incompatible with primary (2) . This absurdity establishes $(9)^2(RU)^1$.

Primary (6) reduces to the inside option $(0, 1)^1$. Since $(1)^3$ and $(7)^3$ via "1" yields $(0)^3$, and $(1)^2$ also yields $(0)^3$, therefore $(0)^3$.

Try $(B)^2$. Then $(B)^2(RS)^3$ via "X" gives $(X)^2$, and $(B)^2(2)^1$ via "B" gives $(G)^2$, and $(G)^2(F^*)^5$ via "F" gives $(F)^2$. Then $(FTX)^4$ via "X" gives $(D)^4$. But $(BDFGX)^4$ is incompatible with primary (2) . Trial $(B)^2$ fails and therefore $(B)^3$.

In conclusion $(BEV)^3$ via "X" implies $(5)^3$, and $(BB^*)^3$ via "B" implies $(G)^3$. Then $(D^*)^2(G)^3$ via "D" gives $(1)^2$, so that the option $(0, 1)^1$ gives $(0)^1$. Now $(0)^1(1)^2$ via "1" gives $(7)^1(4)^2$, and $(0)^1(9)^2$ via "3" gives $(3)^1$. Again $(05)^3(8)^4$ via "2" gives $(8)^5$. But $(C^*F^*0378)^3$ is incompatible with primary (8) . Thus case 12 falls.

All ramifications of the combinatorial problem have been investigated, and we may conclude that there are six and only six solutions, as indicated.

A NOTE ON FINITELY-ADDITIVE MEASURES.*

By ELLEN F. BUCK and R. C. BUCK

1. Introduction. In a previous paper, one of the authors has discussed a finitely-additive measure defined in I , the positive integers.¹ In Section 1, we show that this is typical in that any separable measure in a countable space is equivalent to this measure under a 1:1 map of the space onto the integers.² In Section 2, we define continuous maps of I into the interval $(0, 1)$ which are non-trivial and which generate there the usual measure. Section 3 and 4 deal with certain aspects of the topology defined by the arithmetic progressions.

2. A mapping theorem. Consider a countable class \mathcal{D}_0 of subsets of I , the space of positive integers. \mathcal{D}_0 is the minimal class containing all finite sets and arithmetic progressions, which is closed under finite union and difference. In \mathcal{D}_0 a finitely-additive measure Δ may be defined such that $\Delta(\{an + b\}) = 1/a$ while $\Delta(F) = 0$ if F is finite. Let \mathcal{D}_0^* be the finite Carathéodory closure $\text{cl}[\mathcal{D}_0, \Delta]$ of \mathcal{D}_0 with respect to Δ . A set S belongs to \mathcal{D}_0^* and has measure $\Delta^*(S)$ if for any $\epsilon > 0$, there exist sets $B_1, B_2 \in \mathcal{D}_0$ with $B_1 \subseteq S \subseteq B_2$ and $\Delta(B_2 - B_1) < \epsilon$.

Let X be a countable space in which there is a finitely-additive measure Ω defined on a class of subsets \mathcal{M} , closed under finite union and difference, such that the measure of a point is zero, and $\Omega(X) = 1$. If this measure $[\mathcal{M}, \Omega]$ is in addition *separable*, so that there is a countable subclass \mathcal{M}_0 of \mathcal{M} satisfying: (i) $\mathcal{M} = \text{cl}[\mathcal{M}_0, \Omega]$, (ii) \mathcal{M}_0 contains no infinite sets of measure zero; and, if, (iii) $\Omega(A)$ is rational for every set of \mathcal{M}_0 , then the following theorem holds.

THEOREM 1. *There exists a 1:1 map T of X onto I such that $T\{\mathcal{M}_0\} \subseteq \mathcal{D}_0$, $T\{\mathcal{M}\} \subseteq \mathcal{D}_0^*$, and if $A \in \mathcal{M}$, then $\Omega(A) = \Delta^*(T(A))$.*

* Received November 1, 1946.

¹ R. C. Buck, "The measure theoretic approach to density," *American Journal of Mathematics*, vol 68 (1946), pp. 560-580. This paper will be henceforth referred to as [M].

² This includes an unpublished result on sets having density, due, we believe, to S. Ulam.

We may clearly assume that \mathcal{M}_0 contains X and all the finite subsets of X , and is closed under union and difference; for if not, we adjoin these sets and take the finite Borel extension which is still countable. Restriction (iii) is relatively unimportant. Since the measure values which occur for sets in \mathcal{D}_0^* are all the values in $[0, 1]$, the theorem, without (iii), holds if $T\{\mathcal{M}_0\} \subseteq \mathcal{D}_0$ is deleted. Condition (iii) is clearly necessary for the stronger form. We shall denote the cardinal number of any set S by $|S|$.

LEMMA. Let $A \in \mathcal{M}_0$, $B \in \mathcal{D}_0$ be sets such that $|A| = |B|$, $\Omega(A) = \Delta(B)$. If $A_0 \subseteq A$, $A_0 \in \mathcal{M}_0$, then there exists a set $B_0 \subseteq B$ belonging to \mathcal{D}_0 such that: (a) $\Omega(A_0) = \Delta(B_0)$, (b) $|A_0| = |B_0|$, (c) $|A - A_0| = |B - B_0|$.

Proof. (i) Suppose that $0 < \Omega(A_0) = p/q < \Omega(A) = r/s$. Write $B = \bigcup_{i=1}^{qr} P_i \cup F$ where P_i are disjoint progressions with common difference qs , and F is finite. Choose ps of the P_i in order of their first terms and call their union B_0 . We see that $B_0 \subseteq B$, $B_0 \in \mathcal{D}_0$, and $\Delta(B_0) = p/q = \Omega(A_0)$. Moreover, B_0 is infinite and $B - B_0$ consists of $qr - ps > 0$ progressions P_i and is also infinite. But so were A_0 and $A - A_0$. Thus (b) and (c) hold.

(ii) Suppose $\Omega(A_0) = 0$. Then A_0 is finite and $|A_0| + |A - A_0| = |A|$. Choose B_0 as the first $|A_0|$ terms of B .

(iii) Suppose $\Omega(A_0) = \Omega(A)$. Choose $B - B_0$ to be the first $|A - A_0|$ terms of B .

Consider the vector $\epsilon = (\epsilon_1, \epsilon_2, \dots, \epsilon_n)$ where ϵ_i is 0 or 1. Set $\bar{\epsilon} = \sum \epsilon_i$ and for any set S define $\epsilon_i S$ to be S if $\epsilon_i = 1$, and to be its complement S' if $\epsilon_i = 0$. Let $\epsilon(A_1, A_2, \dots, A_n)$ be $\bigcap_{i=1}^n \epsilon_i A_i$. Thus if $\epsilon \neq \epsilon'$, $\epsilon'(A_1, \dots, A_n) \cap \epsilon(A_1, \dots, A_n) = 0$ and $\bigcup_{\bar{\epsilon} > 0} \epsilon(A_1, \dots, A_n) = \bigcup_1^n A_i$.

Order the sets of \mathcal{M}_0 into a sequence $\{A_i\}$, ($i = 1, 2, \dots$). Since $A_1 \subseteq X$, $\Omega(X) = 1 = \Delta(I)$ and $|X| = |I|$, we can choose, by the lemma, a set $B_1 \in \mathcal{D}_0$ such that $\Delta(B_1) = \Omega(A_1)$ and $|B_1| = |A_1|$, $|X - A_1| = |I - B_1|$. There exists a 1:1 map T of A_1 onto B_1 .

We extend this map to $A_1 \cup A_2$. Consider $A_2 - A_1 \in \mathcal{M}_0$; it is contained in $X - A_1 \in \mathcal{M}_0$ and, applying the lemma, there exists a set $E_1 \in \mathcal{D}_0$, $E_1 \subseteq I - B_1$, such that $|A_2 - A_1| = |E_1|$,

$$|(X - A_1) - (A_2 - A_1)| = |(I - B_1) - E_1|$$

and $\Omega(A_2 - A_1) = \Delta(E_1)$. Let τ be a 1:1 map of $A_2 - A_1$ onto E_1 . Consider now $A_2 \cap A_1$; again applying the lemma, there exists a set $E_2 \in \mathcal{D}_0$,

$E_2 \subseteq B_1$ with $|E_2| = |A_2 \circ A_1|$, $|B_1 - E_2| = |A_1 - (A_1 \circ A_2)|$ and $\Delta(E_2) = \Omega(A_2 \circ A_1)$. Since T is a 1:1 map, $T(A_2 \circ A_1)$ is a subset of B_1 with the same cardinal number relations as $A_2 \circ A_1$, and therefore as E_2 . Let σ be a 1:1 map of B_1 onto itself which maps $T(A_2 \circ A_1)$ onto E_2 . Set $B_2 = E_1 \cup E_2$.

Define the extended T to be τ on $A_2 - A_1$, and σT on A_1 . T is now defined on $A_1 \cup A_2$. It may be easily verified that T has the following properties for $n = 2$.

$$(\alpha) \quad T(A_i) = B_i, \quad |A_i| = |B_i|, \quad \Omega(A_i) = \Delta(B_i), \quad (i = 1, 2, \dots, n).$$

$$(\beta) \quad T(\epsilon(A_1, A_2, \dots, A_n)) = \epsilon(B_1, B_2, \dots, B_n); \quad \text{all } \bar{\epsilon} > 0.$$

$$(\gamma) \quad |\epsilon(A_1, A_2, \dots, A_n)| = |\epsilon(B_1, B_2, \dots, B_n)| \\ \Omega(\epsilon(A_1, A_2, \dots, A_n)) = \Delta(\epsilon(B_1, B_2, \dots, B_n)), \quad \text{all } \epsilon.$$

Suppose that T has been defined on $S_k = \bigcup_{i=1}^k A_i$ with properties (α) , (β) , and (γ) holding for $n = k$. We extend T to $S_k \cup A_{k+1} = S_{k+1}$. Let $F_0 = A_{k+1} - S_k$. On this set T has not yet been defined. (β) and (γ) imply that $T(S_k) = \bigcup_{i=1}^k B_i$, $|S_k| = |\bigcup_{i=1}^k B_i|$, $\Omega(S_k) = \Delta(\bigcup_{i=1}^k B_i)$, and $|X - S_k| = |I - \bigcup_{i=1}^k B_i|$. Thus, by the lemma, there exists a set $E_0 \subseteq I - \bigcup_{i=1}^k B_i$ such that $E_0 \in \mathcal{D}_0$, $|E_0| = |F_0|$, $|(X - S_k) - F_0| = |(I - \bigcup_{i=1}^k B_i) - E_0|$ and $\Omega(F_0) = \Delta(E_0)$. Let τ be a 1:1 map of F_0 onto E_0 .

Let $F_\epsilon = A_{k+1} \circ \epsilon(A_1, A_2, \dots, A_k)$, $\bar{\epsilon} > 0$. $T(F_\epsilon)$ has already been defined; however, by (γ) and the lemma, there is a set E_ϵ in $\epsilon(B_1, \dots, B_k)$ such that $\Delta(E_\epsilon) = \Omega(F_\epsilon)$, $|E_\epsilon| = |F_\epsilon| = |T(F_\epsilon)|$, and

$$|\epsilon(B_1, \dots, B_k) - E_\epsilon| = |\epsilon(A_1, \dots, A_k) - F_\epsilon| = |\epsilon(B_1, \dots, B_k) - T(F_\epsilon)|.$$

Let σ_ϵ be a 1:1 map of $\epsilon(B_1, \dots, B_k)$ onto itself such that $T(F_\epsilon)$ is mapped onto E_ϵ . Set $B_{k+1} = E_0 \cup \bigcup_{\bar{\epsilon} > 0} E_\epsilon$ and define T^* , the extended T , on S_{k+1} to be τ on $A_{k+1} - S_k$ and $\sigma_\epsilon T$ on $\epsilon(A_1, \dots, A_k)$. T^* has properties (α) , (β) and (γ) for $n = k + 1$. We give here only the proof of (α) ; the others may be proved in a similar fashion. Suppose that $i \leq k$:

$$\begin{aligned} T^*(A_i) &= T^*\left(\bigcup_{\epsilon_i=1} \epsilon(A_1, \dots, A_k)\right) = \bigcup_{\epsilon_i=1} \sigma_\epsilon T(\epsilon(A_1, \dots, A_k)) \\ &= \bigcup_{\epsilon_i=1} \sigma_\epsilon(\epsilon(B_1, \dots, B_k)) = \bigcup_{\epsilon_i=1} \epsilon(B_1, \dots, B_k) = B_i. \end{aligned}$$

Thus, none of the sets B_i , $i \leq k$, have been altered and $|A_i| = |B_i|$, $\Omega(A_i) = \Delta(B_i)$. If $i = k + 1$, then

$$\begin{aligned}
T^*(A_{k+1}) &= T^*((A_{k+1} - S_k) \cup (A_{k+1} \cap S_k)) \\
&= \tau(F_0) \cup T^*(A_{k+1} \cap \bigcup_{\epsilon > 0} \epsilon(A_1, \dots, A_k)) \\
&= E_0 \cup \bigcup_{\epsilon > 0} \sigma_\epsilon T(F_\epsilon) = E_0 \cup \bigcup_{\epsilon > 0} E_\epsilon = B_{k+1}.
\end{aligned}$$

Since T^* is 1 : 1, $|A_{k+1}| = |B_{k+1}|$. Finally,

$$\begin{aligned}
\Delta(T^*(A_{k+1})) &= \Delta(E_0 \cup \bigcup_{\epsilon > 0} E_\epsilon) = \Delta(E_0) + \sum_{\epsilon > 0} \Delta(E_\epsilon) \\
&= \Omega(F_0) + \sum_{\epsilon > 0} \Omega(F_\epsilon) = \Omega(A_{k+1}).
\end{aligned}$$

We have thus defined the mapping T recursively; since every point p of X occurs in a unit class $\{p\}$ in \mathcal{M}_0 , $T(p)$ is determined and its image will not be altered in further stages of the induction. Moreover, since X occurs in \mathcal{M}_0 , X is mapped onto I by T , and (α) , (β) , (γ) hold for all n , proving the first part of the theorem.

Now, suppose that $R \in \mathcal{M}$; since \mathcal{M} is the finite Carathéodory extension of \mathcal{M}_0 , for any $\epsilon > 0$, we can find sets A_i and A_j of \mathcal{M}_0 such that $A_i \subset R \subset A_j$, and $\Omega(A_i - A_j) < \epsilon$. Let $T(R) = S$. Then $T(A_i) = B_i \subset S \subset B_j$ and $\Delta(B_i - B_j) < \epsilon$ so that $S \in \mathcal{D}^*_0$, and since $|\Omega(R) - \Delta^*(S)| < \epsilon$ for any $\epsilon > 0$, $\Omega(R) = \Delta^*(S)$. This completes the proof of the theorem.

More generally, let $[\mathcal{M}, \Omega]$ be any finitely-additive measure on a space X , and let ϕ be a 1 : 1 map of X onto X . Let $\mathcal{M}^* = \phi\{\mathcal{M}\}$, and define $\Omega^*(A)$, for $A \in \mathcal{M}^*$, to be $\Omega\phi^{-1}(A)$. Then, $[\mathcal{M}^*, \Omega^*]$ is also a finitely-additive measure on X . We call any measure obtained in this fashion a replica of $[\mathcal{M}, \Omega]$. It is clear that any replica of a separable measure is separable.

COROLLARY. Any separable measure on I is a contraction of a replica of $[\mathcal{D}^*_0, \Delta^*]$.

For, given $[\mathcal{M}, \Omega]$ on I , there is a map T of I onto I such that $T\{\mathcal{M}\} \subseteq \mathcal{D}^*_0$, and $\Omega(A) = \Delta^*T(A)$. If $[\mathcal{M}^*, \Omega^*]$ is the replica of $[\mathcal{D}^*_0, \Delta^*]$ given by $\phi = T^{-1}$, then $\mathcal{M} \subseteq \mathcal{M}^*$ and $\Omega^*(A) = \Omega(A)$ for $A \in \mathcal{M}$.

3. Continuous functions. Let Y and Z be topological spaces and f be a continuous mapping from Y into Z . Then if a measure m is defined on open sets of Y , a measure m^* can be defined for open sets O of Z by $m^*(O) = mf^{-1}(O)$. It is clearly sufficient that m need be defined only on open sets which arise as inverse images of open sets of Z . We shall discuss this construction when $Y = I$, $Z = (0, 1)$, the open unit interval, and $m = \Delta^*$; we

shall show that for a simply defined function f , m^* coincides with Euclidean length for open intervals (α, β) .

We introduce a natural topology into I by choosing for neighborhoods all progressions. Thus a general neighborhood of a point $p \in I$ will be the progression $\{\lambda n + p\}$, ($n = 0, 1, 2, \dots$). These sets are both open and closed and so the space is totally disconnected.³ We may define many continuous real-valued functions. For example, the characteristic function of any neighborhood is continuous. Such a function is, in a sense, a trivially continuous function even though it is not constant. For our purpose, we construct a non-trivial 1:1 real-valued continuous function.

For any $N \in I$, we may express N uniquely in the form $N = \alpha_0 + \alpha_1 r + \alpha_2 r^2 + \dots + \alpha_n r^n$ where r and α_i are integers, $r \geq 2$ and $0 \leq \alpha_i < r$. Let $f_r(N) = \alpha_0 r^{-1} + \alpha_1 r^{-2} + \dots + \alpha_n r^{-n-1}$. This is uniquely defined for all N and maps I in a 1:1 manner onto the set of rational numbers in $(0, 1)$ of the form A/r^k .

THEOREM 2. *The function f_r is continuous, and if O is the open interval (α, β) on $(0, 1)$, then $f_r^{-1}(O) \in \mathcal{D}_0^*$ and $\Delta^* f_r^{-1}(O) = \beta - \alpha$.*

Proof. If $f_r(N) \in (Ar^{-k}, (A+1)r^{-k})$, then $f_r(N) = Ar^{-k} + \alpha_k r^{-k-1} + \alpha_{k+1} r^{-k-2} + \dots$ where the α_i are ultimately zero but at least one is not zero. Write $Ar^{-k} = \alpha_0 r^{-1} + \alpha_1 r^{-2} + \dots + \alpha_{k-1} r^{-k}$; then

$$N = (\alpha_0 + \alpha_1 r + \dots + \alpha_{k-1} r^{k-1}) + r^k (\alpha_k + \alpha_{k+1} r + \dots)$$

and $N = a + r^k n$ where n is some positive integer. As $f_r(N)$ ranges over $(Ar^{-k}, (A+1)r^{-k})$, n ranges over all positive integers and thus $f_r^{-1}\{(Ar^{-k}, (A+1)r^{-k})\}$ is the progression $\{r^k n + a + r^k\}$ ($n = 0, 1, 2, \dots$), which is an open set. Similarly, the inverse image of the interval $[Ar^{-k}, (A+1)r^{-k})$ is the progression $\{r^k n + a\}$ ($n = 0, 1, 2, \dots$). Thus the inverse image of (Ar^{-k}, Br^{-k}) is the sum of $B - A$ disjoint progressions of the form $\{r^k n + a\}$, is open and belongs to \mathcal{D}_0 . Hence f_r is continuous. Moreover, $\Delta f_r^{-1}\{(Ar^{-k}, Br^{-k})\} = r^{-k}(B - A)$.

For any (α, β) in $(0, 1)$, we may choose rational numbers such that for large k ,

³ This implies that the space is metrizable; for example, let

$$\rho(x, y) = \sum_{a, b=1}^{\infty} |f_{a,b}(x) - f_{a,b}(y)| 2^{-a-b}$$

where $f_{a,b}$ is the characteristic function of $\{an + b\}$. The space is not compact. Topologies of this type on the additive group of integers are often used as illustrations—e. g., Pontrjagin, *Topological Groups*, p. 57.

$$A_k r^{-k} \leq \alpha \leq A'_k r^{-k} < B'_k r^{-k} \leq \beta \leq B_k r^{-k}$$

and

$$\lim A_k r^{-k} = \lim A'_k r^{-k} = \alpha, \quad \lim B_k r^{-k} = \lim B'_k r^{-k} = \beta.$$

If $S = f_r^{-1}\{(\alpha, \beta)\}$, the sets of \mathcal{D}_0 corresponding to $(A'_k r^{-k}, B'_k r^{-k})$ and $(A_k r^{-k}, B_k r^{-k})$ approximate S from within and without such that the Δ -measure of their difference is $(B_k - A_k) r^{-k} - (B'_k - A'_k) r^{-k}$ which approaches zero. Thus the set S belongs to \mathcal{D}^*_0 and $\Delta^*(S) = \lim (B_k - A_k) r^{-k}$ or $\Delta^*(S) = \beta - \alpha$.

4. Decompositions. We cannot conclude that the inverse image of an arbitrary open set in $(0, 1)$ is measurable Δ^* for there are open sets in I which do not belong to \mathcal{D}^*_0 . Any open set in I is a union of progressions; since the union and difference of any two progressions is either null or a finite union of disjoint progressions, we may express any open set as a countable union of *disjoint* progressions. Let $\mathcal{D} \supset \mathcal{D}^*_0$ be the class of subsets of I for which a density exists, defined as $D(S) = \lim S(N)/N$ where $S(N)$ is the number of terms of S not greater than N .

THEOREM 3. *Let S be an open set in I , $S = \bigcup B_i$, where $B_i = \{b_i n + a_i\}$ and $B_i \cap B_j = 0$ if $i \neq j$. Let $A = \{a_i\}$. Then, $S - A \in \mathcal{D}$ and $D(S - A) = \beta = \Sigma \Delta(B_i)$.*

Proof. Clearly $S(N) = \Sigma B_i(N)$. Since $B_i = \{b_i n + a_i\}$,

$$B_i(N) = \begin{cases} 1 + [(N - a_i)/b_i] & \text{if } a_i \leq N \\ 0 & \text{if } a_i > N. \end{cases}$$

Thus

$$S(N) = \sum_{a_i \leq N} 1 + \sum_{a_i \leq N} [(N - a_i)/b_i]$$

or

$$(S(N) - A(N))/N = 1/N \sum_{a_i \leq N} [(N - a_i)/b_i] = u_N.$$

Since

$$u_N \leq 1/N \sum_{a_i \leq N} N/b_i = \sum_{a_i \leq N} 1/b_i \leq \beta,$$

$\lim u_N \leq \beta$. For the other direction, choose some index k ; if N is sufficiently large so that $N > a_k$,

$$u_N \geq \sum_{a_i \leq a_k} [(N - a_i)/b_i]/N$$

and as N increases,

$$\lim u_N \geq \sum_{a_i \leq a_k} 1/b_i. \quad \text{Then letting } k \text{ increase, } \lim u_N \geq \beta.$$

COROLLARY 1. *$S \in \mathcal{D}$ if and only if $A \in \mathcal{D}$ and then $D(S) = D(A) + \beta$.*

Since I is open, we can express it as the union of disjoint progressions; in particular, $I = \bigcup_{\lambda=1}^{\infty} 2^{\lambda}n + 2^{\lambda-1}$. More generally, if r is any integer, $r \geq 2$, $I = \bigcup B_i$ where $B_i = \{r^i n + a_i\}$, and $B_i \cap B_j = 0$ if $i \neq j$. To effect this we need only choose $a_1 = 1$, and choose $a_{\lambda+1}$ as the least integer in $I - \bigcup_{i=1}^{\lambda} B_i$. From this decomposition we see at once that there are a non-countable number of sets both open and closed. Observe also that

$$\beta = \sum \Delta(B_i) = \sum 1/r^i = 1/(r-1)$$

which may be made arbitrarily small. Applying the above theorem to such a decomposition, we obtain in particular:

COROLLARY 2. *There exist open sets not in \mathcal{D} .*

Let $I = \bigcup \{3^i n + a_i\} = \bigcup B_i$, $B_i \cap B_j = 0$. Since $I \in \mathcal{D}$, $A = \{a_i\}$ belongs to \mathcal{D} and $D(A) = 1 - \beta = 1/2 > 0$. Select a subset A_0 of A which does not have a density, and define S as $\bigcup_{a_i \in A_0} B_i$. By Corollary 1, this open set cannot belong to \mathcal{D} .

Such decompositions of I also yield open sets which belong to \mathcal{D} but not to \mathcal{D}_0^* . A set E is everywhere dense if its closure is I ; then, every progression meets E in an infinite number of points.⁴ This is equivalent to $\bar{\Delta}(E) = 1$ where $\bar{\Delta}(E) = \inf \Delta(B)$ for $E \subseteq B$ and B in \mathcal{D}_0 . [M].

THEOREM 4. *If $I = \bigcup \{r^i n + a_i\}$ is a disjoint decomposition of I , $A = \{a_i\}$, then $A' = I - A$ is an open everywhere dense set belonging to \mathcal{D} , but if $r > 2$, not to \mathcal{D}_0^* .*

Proof. $A' = \bigcup \{r^i n + a_i + r^i\}$ and is open but not closed. To show that A' is everywhere dense it is sufficient to show that every progression $\{\lambda m + a_j\}$ meets A' infinitely often for every λ and j ($\lambda, j = 1, 2, \dots$). For then the closure of A' is I . If two progressions have one point in common, their intersection is infinite. If we choose $i = j$, $m = r^j$, $n + 1 = \lambda \geq 1$, then for any λ and j , $\lambda m + a_j = r^i n + a_i + r^i$; thus $\bar{\Delta}(A') = 1$. By the previous theorem, A' belongs to \mathcal{D} and $D(A) = (r-2)/(r-1)$. Since D agrees with Δ on \mathcal{D}_0 , $\bar{\Delta}(A) \geq D(A)$. If $C_k = \bigcup_{i=1}^k \{r^i n + a_i + r^i\}$, then $A \subset C'_k$ and

⁴ Added in proof. Everywhere dense sets predominate. A measure may be defined in $2I$ by a dyadic mapping of this onto $(0, 1)$. [M] By a standard application of Rademacher functions, one may show the following: if $A \subset I$ does not have unit lower density, the class of subsets of A has zero measure. Hence, almost every subset of I is everywhere dense in I (extremal) and \mathcal{D}_0^* has measure zero. This answers a question raised in [M, p. 580].

$$\bar{\Delta}(A) \leq \lim \Delta(C'_k) = 1 - \sum_1^{\infty} 1/r^k.$$

Hence $\bar{\Delta}(A) = (r-2)/(r-1)$. If $r > 2$, $\bar{\Delta}(A) + \bar{\Delta}(A') > 1$ and A' is non-measurable. We remark that A is nowhere dense in I but $D(A)$ may be arbitrarily close to $D(I) = 1$.

5. Special sets. Since the topology is inherently arithmetic, it is natural to look for connections with the theory of numbers.

THEOREM 5. *The set of primes, P , is nowhere dense, has 1 as its only limit point but is not convergent.*

Since $\Delta^*(P) = 0$, the set of composites $C = I - P - \{1\}$ is everywhere dense. [M]. But $C = \bigcup_{p \in P} \{np + p\}$ and so C is open, $P \cup \{1\}$ is closed and therefore nowhere dense. P has 1 as a limit point if and only if every progression $\{\lambda n + 1\}$ contains an infinite number of primes; this follows from Dirichlet's theorem.⁵ For any $q \neq 1$, the neighborhood $\{qn + q\}$ contains at most one prime and q is not a limit point of P . Since $\{4n + 3\}$ contains an infinite number of primes but does not contain 1, P does not converge. We can, of course, choose a subset of the primes convergent to 1. For example, let p_k be the least prime in $\{(k+1)!n + 1\}$ distinct from p_1, p_2, \dots, p_{k-1} .

THEOREM 6. *The set $\{m^2\}$ is closed, perfect, and nowhere dense.*

Let q be a limit point of $\{m^2\}$. Then for any λ , $\lambda n + q$ must belong to $\{m^2\}$ for an infinite number of n . Thus, $x^2 - q \equiv 0 \pmod{\lambda}$ has a solution for each λ . It then follows that q is a square. If $p = \alpha^2$, then for each λ , and for all t , $(\alpha + \lambda t)^2 = \alpha^2 + \lambda(t^2\lambda + 2t\alpha)$. Hence $\{\alpha^2 + \lambda n\}$ intersects $\{m^2\}$ infinitely often. Thus $\{m^2\}$ is closed and perfect. Moreover, $\Delta^*(\{m^2\}) = 0$ and so $\{m^2\}$ is nowhere dense. [M].

THEOREM 7. *$S = \{[\alpha n + \beta]\}$ where α is larger than 1 and irrational, $\beta \geq 0$, is neither open nor closed, and is everywhere dense.*

The proof follows immediately since $\bar{\Delta}^o(S) = 1 = \bar{\Delta}(S')$. [M]. This result is dual to the classical one which states that if α is irrational, the fractional parts of $\alpha n + \beta$ are everywhere dense in $(0, 1)$.

WELLESLEY COLLEGE AND
SOCIETY OF FELLOWS, HARVARD UNIVERSITY.

⁵ Landau, *Vorlesungen über Zahlentheorie*, vol. 1. An elementary proof of the case used here can also be given. G. D. Birkhoff and H. S. Vandiver, "On the integral divisors of $a^n - b^n$," *Annals of Mathematics*, Ser. 2, vol. 5 (1903), pp. 173-180.

ON CONFORMAL CORRESPONDENCE OF SURFACES AND MANIFOLDS.*

By P. SAMUEL.

I. Introduction.

In the ordinary euclidean 3-space the problem of finding two surfaces in conformal correspondence with parallel corresponding tangent planes was solved by Christoffel.¹ The method is the consideration of the linear correspondence between the corresponding vectors of the tangent planes: one takes the principal directions of this linear operator as coordinate directions. The result is: either the two surfaces are minimal, or their lines of curvature form an isothermic net, according as the principal directions are isotropic (the linear correspondence being a proper similarity) or real (the linear correspondence being a proper similarity followed by a symmetry).

This method of taking the principal directions of a suitable linear operator as coordinate directions is very often used in differential geometry: for example the consideration of the lines of curvature.

In this paper we shall study the general case of two n -manifolds in an euclidean N -space, in conformal correspondence, the tangent n -planes at two corresponding points being parallel. We shall use the same general method: take the principal directions of the linear operator between the corresponding vectors in the tangent n -planes as coordinate directions. For convenience this operator will be called "parallel tangent planes operator" or "the operator."

After having fixed the notations and the terminology used, I establish in II and III some properties of the operator which will be used throughout this paper. IV is a rapid review of some important kinds of manifolds in the euclidean N -space. V studies the case of 2-dimensional manifolds; the results obtained are a mere generalization of those of Christoffel: we obtain as only solutions the minimal surfaces and the surfaces which have isothermic lines of curvature. With manifolds of any dimension the case where all the principal directions of the operator are real can be studied completely; this

* Received December 9, 1946.

¹ Christoffel, E. B., "Über einige allgemeine Eigenschaften der Minimumsflächen," *Journal für die reine und angewandte Mathematik*, vol. 67 (1867), p. 218. See also: Darboux, "Théorie des surfaces," t. II, livre IV, ch. XI, pp. 239-55, 1889.

is done in VI; the result is that the manifolds must belong to a very restricted class; an application is given to the hyperspherical representation of a hypersurface: when the space is of dimension > 3 it is only conformal for the hypersphere itself and the generalization of the catenoid. In VII is studied the case of 3-dimensional manifolds: if the principal directions of the operator are holonomic we give a complete solution, that is parametric representations of the two manifolds in terms of 3 arbitrary vector functions of one variable; in the non-holonomic case we get only the ds^2 of the manifolds in terms of an arbitrary function of two variables and some necessary and sufficient conditions on the second quadratic forms. VIII is the study of the general case, with the restriction that the principal directions are holonomic; a complete answer is given: the only possible manifolds are essentially: the translation manifolds generated by two totally isotropic manifolds and their cartesian products,—and a generalization of the manifolds obtained in VII. IX is the study of the case where the correspondence between the two manifolds is not only conformal but isometric; using classical lemmas about geodesics and geodesic coordinates, we prove that the linear correspondence between the two parallel tangent n -planes (which is a rotation) has its angles constant when decomposed in its canonical form; after that we consider the case of 3- and 4-dimensional manifolds, and we prove the holonomy of the principal directions; we are therefore reduced to the case studied in the preceding section.

It should be noted that our considerations are only *local* and restricted to *analytic manifolds*.

Notations and terminology.

In this paper we use the notations and the terminology of J. A. Schouten.² $\partial_a x$ denotes the ordinary derivative with respect to the variable u^a , or the non-holonomic derivative with respect to the coordinate direction of index a ; $\Gamma^\gamma_{\alpha\beta}$ are the Christoffel symbols defining the affine connection of the manifold considered as a Riemannian space; $g_{\alpha\beta}$ is the metric tensor; ∇_a is the covariant derivation with respect to the affine connection; $\partial_{\alpha\beta}$ and $\nabla_{\alpha\beta}$ denote the second derivatives in the order β, α . $\Omega^\gamma_{\alpha\beta}$ is the "non-holonomy object" relative to our coordinate directions, which are in the general case non-holonomic; this symbol expresses the "deviation of the commutativity" of the second derivatives in terms of the first ones $\partial_{\alpha\beta} - \partial_{\beta\alpha} = \Omega^\gamma_{\beta\alpha} \partial_\gamma$. We use the classical summation convention.

² Schouten, J. A., "*Einführung in die neueren Methoden der Differentialgeometrie*," Groningen, 1935.

The corresponding points on the two manifolds will be denoted by M and P , which will denote also the vectors \vec{OM} and \vec{OP} , O being the origin in the euclidean N -space. The two manifolds will be represented in terms of the same (holonomic or not) parameters u^a , two corresponding points being obtained for the same values of the parameters, as usual. Some quantities, like the $\Omega^{\gamma}_{\alpha\beta}$, are the same for the two manifolds; for the others we must use a subscript: $g_{M,\alpha\beta}$, $g_{P,\alpha\beta}$, $\Gamma_M^{\gamma}_{\alpha\beta}$, $\Gamma_P^{\gamma}_{\alpha\beta}$.

II. General Properties of the Operator.

We denote by $\{\vec{n}_a\}$ a free system of $N - n$ vectors orthogonal to the tangent n -planes at M (or P). The Frenet-formulae of the manifold (M) are:

$$\partial_{\alpha\beta}M = \Gamma_M^{\gamma}_{\alpha\beta} \cdot \partial_{\gamma}M + a_{M\alpha\beta}^a \cdot \vec{n}_a$$

(The $a_{\alpha\beta}^a du^a du^{\beta}$ are the second fundamental forms of the manifold (M) imbedded in the euclidean N -space.)

We define the operator b_a^{β} by the relation $\partial_a P = b_a^{\beta} \cdot \partial_{\beta} M$. If the tensor b_a^{β} is in diagonal form, the coordinate directions being its principal directions, then $b_a^{\beta} = \delta_a^{\beta}$, and $\partial_a P = \partial_a M$ (α not summed).

We consider the compatibility conditions

$$\begin{aligned} \partial_{\beta\alpha}P - \partial_{\alpha\beta}P &= \Omega^{\gamma}_{\alpha\beta} \cdot \partial_{\gamma}P = \gamma s \cdot \Omega^{\gamma}_{\alpha\beta} \cdot \partial_{\gamma}M = \partial_{\beta}(\gamma s) \cdot \partial_{\alpha}M - \partial_{\alpha}(\gamma s) \cdot \partial_{\beta}M \\ &+ \gamma s \cdot \partial_{\beta\alpha}M - \gamma s \cdot \partial_{\alpha\beta}M. \end{aligned}$$

Or

$$\begin{aligned} \gamma s \cdot \Omega^{\gamma}_{\alpha\beta} \cdot \partial_{\gamma}M &= \partial_{\beta}(\gamma s) \cdot \partial_{\alpha}M - \partial_{\alpha}(\gamma s) \cdot \partial_{\beta}M \\ &+ (\gamma s \Gamma_{M\beta\alpha}^{\gamma} - \gamma s \cdot \Gamma_{M\alpha\beta}^{\gamma}) \partial_{\gamma}M + (\gamma s - \gamma s) a_{M\alpha\beta}^a n_a. \end{aligned}$$

From these vector equalities we conclude

$$a) \quad (\gamma s - \gamma s) a_{M\alpha\beta}^a = 0.$$

Two principal directions relative to two distinct characteristic roots are conjugate with respect to all the second forms.

b) If γ is different from α and β

$$(1) \quad \gamma s \cdot \Omega^{\gamma}_{\alpha\beta} = \alpha s \cdot \Gamma_M^{\gamma}_{\beta\alpha} - \beta s \cdot \Gamma_M^{\gamma}_{\alpha\beta}.$$

c) If $\gamma = \alpha$ (same formula if $\gamma = \beta$)

$$(2) \quad \alpha s \cdot \Omega_{(\alpha)\beta}^{(\alpha)} = \partial_{\beta}(\alpha s) + \alpha s \cdot \Gamma_{M\beta(\alpha)}^{(\alpha)} - \beta s \cdot \Gamma_{M(\alpha)\beta}^{(\alpha)}.$$

The parenthesis (α) denotes that there is no summation with respect to α .

III. Principal Directions when the Correspondence is Conformal.

In this case $ds_P^2 = A^2 ds_M^2$. Or $g_{Pa\beta} = A^2 g_{Ma\beta} = {}_{as}\beta s g_{Ma\beta}$. Suppressing the subscript M we have

$$(3) \quad ({}_{as} \cdot \beta s - A^2) g_{a\beta} = 0.$$

We conclude from (3):

1) If $\alpha = \beta$:

a) $g_{a\beta} = 0$. We obtain $n - p$ isotropic principal directions, complex conjugate by pairs, denoted here by latin indices: $g_{ii} = g_{i\cdot i} = 0$.

b) $g_{aa} \neq 0$ (real principal directions). Then ${}_{as} = \pm A$. The q directions with ${}_{as} = +A$ will be denoted by $\alpha, \beta, \gamma \dots$ (positive directions), and the $p - q$ directions with ${}_{as} = -A$ by $\lambda, \mu, \nu \dots$ (negative directions).

2) For two positive or two negative directions, (3) is verified. For one positive and one negative one must have $g_{a\lambda} = 0$.

3) g_{ii} , sum of products of complex conjugate quantities, is real > 0 ; hence ${}_{is}i \cdot s - A^2 = 0$; ${}_{is} = Ae^{i\theta_i}$; $i \cdot s = Ae^{-i\theta_i}$ ($\theta_i \neq k\pi$).

4) Then one sees easily that $g_{ia} = g_{i\lambda} = 0$, and $g_{ik} = 0$, if $\theta_i + \theta_k \neq 2k\pi$.

IV. Some Classical Manifolds of the Euclidean N -Space.

In this section we shall review a certain number of manifolds imbedded in the euclidean N -space. These manifolds are classical and the only aim of this short review is to give a name to each of them in order to increase the brevity of the rest of this paper.

1) Translation manifolds.

These manifolds can be represented parametrically by

$$M = U(u^a) + V(v^\lambda).$$

The vector U depends only on the variable u^a , V only on the other variables v^λ .

2) Particular case: "cylinder."

In many questions the vectors $\partial_a M$ and $\partial_\lambda M$ must be orthogonal: $\partial_a U \cdot \partial_\lambda V = 0$. They vary independently. If \mathfrak{U} and \mathfrak{B} are the linear subspaces spanned respectively by them, \mathfrak{U} and \mathfrak{B} are totally orthogonal. Then,

up to constant vectors, $U \in \mathfrak{U}$ and $V \in \mathfrak{B}$. In this case it may be useful to take the first axis of cartesian coordinates in \mathfrak{U} , and the last ones in \mathfrak{B} .

3) Revolution manifolds.

In the N -space there are several kinds of revolution manifolds according as the rotation takes place around an $(N-2)$ -plane, an $(N-3)$ -plane, \dots . In the first case, with suitable axis, the parametric representation of the manifold is

$$x_1 = r \cdot \cos \theta; \quad x_2 = r \cdot \sin \theta; \quad x_j = x_j(r), \quad 2 < j \leq N.$$

In the second case

$$x_1 = r \cdot \cos \theta \cdot \cos \phi; \quad x_2 = r \cdot \cos \theta \cdot \sin \phi; \quad x_3 = r \cdot \sin \theta; \quad x_j = x_j(r), \quad 3 < j.$$

4) Totally isotropic manifolds.

A manifold is called totally isotropic if every tangent vector is isotropic (i. e. of length zero). Since $dM = \partial_a M \cdot du^a$, $(dM)^2 = 0$ implies $\partial_a M \cdot \partial_\beta M = 0$: the coordinate directions must be isotropic and mutually orthogonal. This necessary condition (C) is evidently sufficient.

If we take the first n coordinates x^a in the euclidean N -space as parameters on the manifold, the others being denoted by x^a, x^b, \dots the condition (C) is:

$$1 + \sum_{a=1}^{a=N-n} (\partial_a x^a)^2 = 0; \quad \sum_{a=1}^{a=N-n} \partial_a x^a \partial_\beta x^a = 0.$$

If $x^a = iy_a$, these conditions express that the n -manifold of the $(N-n)$ -space, with cartesian coordinates y_a and parameters x^a , has

$$ds^2 = (dx^a)^2 + (dx^\beta)^2 + \dots$$

The study of the totally isotropic manifolds is therefore equivalent to the study of the manifolds with an euclidean ds^2 . Note that we must have $N-n \geq n$, or $n \leq N/2$: in an N -space a totally isotropic manifold is at most of dimension $N/2$. Those of dimension $N/2$ are linear, the manifold with euclidean ds^2 being then the whole $(N-n)$ -space.

5) Translation manifolds generated by two complex conjugate totally isotropic manifolds.

A totally isotropic manifold is always imaginary. Then we consider the translation manifold generated by two complex conjugate totally isotropic manifolds, the real points being obtained by associating two complex conjugate

points: $M = P(u^a) + P^*(u^{a*})$. $g_{a\beta} = g_{a^*\beta^*} = 0$; $g^{a\beta} = g^{a^*\beta^*} = 0$. Since $\partial_{a\beta} M = 0$, $a_{a\beta^*}^i = 0$ for all the second forms; $a_{a\beta^*}^i = g^{a\lambda} a_{\lambda\beta^*}^i + g^{a\lambda^*} a_{\lambda^*\beta^*}^i = 0$, and $a_{a^*\beta^*}^i = 0$. The characteristic equation of the matrix $(a_{a\beta^*}^i)$ is;

$$\begin{vmatrix} s & 0 & . & . \\ 0 & s & . & . \\ . & . & s & 0 \\ . & . & 0 & s \end{vmatrix} = 0.$$

Applying Laplace's rule, we see, two corresponding minors of order $n/2$ having as factor the same power of s , that the characteristic equation will contain only terms with an even power of s . Therefore the roots of the equation, which are the principal curvatures relative to the second form $a_{a\beta^*}^i du^a du^{\beta^*}$, are opposite by pairs; their sum, the mean curvature normal vector, is zero. With two isotropic curves in the 3-space we obtain the minimal surfaces.

V. Two-Dimensional Case.

In this case any net of coordinate lines is holonomic. The two parameters on the surfaces will be called u and v .

1) Two positive principal directions (or two negative ones).

$$\partial_u P = A \partial_u M, \quad \partial_v P = A \partial_v M.$$

The compatibility condition is $\partial_{uv} P - \partial_{vu} P = \partial_u A \cdot \partial_v M - \partial_v A \cdot \partial_u M = 0$. Therefore: $\partial_u A = \partial_v A = 0$, $A = \text{Const}$. The surfaces are homothetic.

2) Two isotropic principal directions.

$$\partial_u P = A e^{i\theta} \cdot \partial_{ii}, \quad \partial_v P = A \cdot e^{-i\theta} \cdot a_v M.$$

The compatibility condition is

$$\partial_v (A e^{i\theta}) \cdot \partial_u M - \partial_u (A e^{-i\theta}) \cdot \partial_v M + 2A i \cdot \sin \theta \cdot \partial_{uv} M = 0.$$

But

$$(\partial_u M)^2 = (\partial_v M)^2 = 0, \quad \partial_u M \cdot \partial_v M \neq 0.$$

By scalar multiplication of the compatibility condition by $\partial_u M$ and $\partial_v M$, and since $\partial_{uv} M \cdot \partial_u M = 1/2 \cdot \partial_v (\partial_u M)^2 = 0$, one obtains $\partial_v (A e^{i\theta}) = \partial_u (A e^{-i\theta}) = 0$. Hence (if $\sin \theta = 0$ we are in the first case) $\partial_{uv} M = 0$. Therefore

$$M = U(u) + V(v), \quad U'^2 = V'^2 = 0.$$

The surfaces are translation surfaces generated by two isotropic curves. Then $P = U_1(u) + V_1(v)$, the curves U and U_1 , V and V_1 having parallel tangents at corresponding points.

3) One positive and one negative principal direction.

These directions are orthogonal (III) and conjugate (II). Therefore the surfaces have lines of curvature, which are taken as coordinate lines.

$$\begin{aligned} ds^2 &= E \cdot du^2 + G \cdot dv^2, & ds^2 &= A(E \cdot du^2 + G \cdot dv^2) \\ \partial_u P &= A \cdot \partial_u M, & \partial_v P &= -A \cdot \partial_v M. \end{aligned}$$

The compatibility condition is

$$2A \cdot \partial_{uv} M + \partial_u A \cdot \partial_v M + \partial_v A \cdot \partial_u M = 0.$$

By scalar multiplication by $\partial_u M$ and $\partial_v M$ we get

$$\begin{aligned} A \cdot \partial_v E + E \cdot \partial_v A &= 0 \quad \text{or} \quad E = f(u)/A \\ A \cdot \partial_u G + G \cdot \partial_u A &= 0 \quad \text{or} \quad G = g(v)/A. \end{aligned}$$

Changing the variables u and v one can take $E = G = 1/A$.

$$ds^2_M = (1/A) \cdot 2(du^2 + dv^2).$$

The lines of curvature are isothermic. Conversely the compatibility condition is verified in this case.

Therefore, to each surface having lines of curvature which are isothermic corresponds another surface with the same property; these two surfaces are in conformal correspondence with parallel tangent planes.

Example: revolution surfaces.

$$x_1 = r \cdot \cos \theta, \quad x_2 = r \cdot \sin \theta, \quad x_j = x_j(r) \quad (j \geq 3).$$

The following vectors are $N-2$ linearly independent normal vectors:

$$(-x'_j \cos \theta, -x'_j \sin \theta, 0, \dots, 0, 1, 0, \dots, 0).$$

The ds^2 is:

$$ds^2 = (1 + \sum_{j \geq 3} x'^2_j) \cdot dr^2 + r^2 \cdot d\theta^2.$$

The $N-2$ second forms relative to the preceding normal vectors are:

$$(x'_j dr^2 + r \cdot x'_j \cdot d\theta^2) / \sqrt{1 + x'^2_j}.$$

They are reduced simultaneously to sums of squares; hence there exist

lines of curvature which are the meridians and the parallels. Since $ds^2 = r^2(du^2 + d\theta^2)$, where $du = \sqrt{1 + \sum_{j \geq 3} x_j'^2(r)} \cdot dr/r$, they are isothermic.

If the coordinates of P are (y_1, y_2, \dots, y_N) , the integration of

$$\partial_r P = 1/r^2 \cdot \partial_r M, \quad \partial_\theta P = -1/r^2 \cdot \partial_\theta M,$$

gives

$$y_1 = -\cos \theta/r, \quad y_2 = -\sin \theta/r, \quad y_j = \int dx_j/r^2 = \int x_j'(r) dr/r^2 \quad (j \geq 3).$$

In order to avoid the integrations one can write

$$x_j = r(\alpha_j - r\alpha_j'), \quad y_j = -(\alpha_j + r\alpha_j')/r$$

α_j being an arbitrary function of r only.

In the case of ordinary space one obtains an interesting correspondence between the meridians, i. e., between plane curves. To the curve $(z(t), r(t))$ corresponds the curve $(\int dz/r, -1/r)$. We take z as the independent variable, and if we denote by R the radius of curvature of the meridian, by N the length of the normal to the meridian limited by the incidence point and the z -axis,

$$R = (1 + r'^2)^{3/2}/r'', \quad N = r(1 + r'^2)^2, \quad R/N = (1 + r'^2)/rr'.$$

For the other curve; $r_1 = -1/r$, $dz_1 = dz/r$.

$$r'_1 = dr_1/dz_1 = \frac{1}{r^2} \frac{dr}{dz} \frac{dz}{dz_1} = (1/r^2)r'r^2 = r' \quad (\text{the tangents are parallel})$$

$$r''_1 = dr'_1/dz_1 = \frac{dr'}{dz} \frac{dz}{dz_1} = r''r^2.$$

$$\text{Hence } R_1/N_1 = -R/N.$$

At corresponding points the quantities R/N are opposite.

To a circle with center on Oz ($N = R$; sphere in the space) corresponds the catenary ($N = -R$; minimal surface of revolution in the space).

To the cycloid with base Oz ($R = 2N$) corresponds the parabola with directrix Oz ($R = -2N$).

There are other interesting couples of curves:

Hyperbola ($z = a \cdot \sinh t$, $r = b \cdot \cosh t$) and sinusoid ($r_1 = -(1/b) \cdot \sin b^2 z_1/a$).

Hyperbola ($z = a \cdot \cosh t$, $r = b \cdot \sinh t$) and ($r = -(1/b) \cdot \sinh b^2 z_1/a$).

Circle tangent to Oz ($z = a \cdot \sin t$, $r = a + \cos t$) and rectifiable cubic of Tschirnhausen

$$(z_1 = -(1/6a)(3u - u^3), \quad r_1 = -(4/6a) - (1/6a)(3u^2 - 1); \\ u = \tan t/2).$$

Parabola with axis Oz ($z = r^2/2p$) and exponential curve ($r_1 = -e^{pz_1}$).

VI. Case where all the Principal Directions are real.

1) The formula of II b) with $\gamma = \lambda$, $\gamma s = -A$, $\alpha s = \beta s = A$, gives $-A \cdot \Omega^\lambda_{\alpha\beta} = A(\Gamma_M^\lambda{}_{\beta\alpha} - \Gamma_M^\lambda{}_{\alpha\beta}) = A\Omega^\lambda_{\alpha\beta}$. Therefore $\Omega^\lambda_{\alpha\beta} = 0$, and for the same reason $\Omega^\alpha_{\lambda\mu} = 0$. Hence the two linear systems $\{\partial_\alpha f = 0\}$ and $\{\partial_\lambda f = 0\}$ are complete. One can find *holonomic coordinate directions* such that the first ones (α) are principal directions relative to A , the others (λ) principal directions relative to $-A$.

2) The formula of II c) gives, if there exist at least two positive directions α, β : $A\Omega^{(\alpha)}_{(\alpha)\beta} = \partial_\beta A + A \cdot \Omega^{(\alpha)}_{(\alpha)\beta}$, or $\partial_\beta A = 0$. In the same way $\partial_\lambda A = 0$.

Therefore A is a constant unless one direction is alone of its kind.

In the case $n = 2$, already studied, the two directions could be of different kinds, and A could depend on the two variables. In the case $n \geq 3$, what is a little surprising at first sight, we have less freedom. If one direction is alone of its kind, the other ones are in number ≥ 2 , and A can depend only on the variable corresponding to the first direction.

Case A constant.

If all the directions are positive (or negative) the two manifolds are homothetic. We suppose now that we have

$$\partial_\alpha P = A \cdot \partial_\alpha M, \quad \partial_\lambda P = -A \cdot \partial_\lambda M.$$

Integrating:

$$P = A(M + \Phi(u^\lambda)), \quad P = A(-M + \Psi(u^\alpha)).$$

Hence

$$P = \frac{1}{2}A(\Psi(u^\alpha) + \Phi(u^\lambda))$$

$$M = \frac{1}{2}(\Psi(u^\alpha) - \Phi(u^\lambda)).$$

But we must have also $\partial_\alpha M \cdot \partial_\lambda M = 0$. Applying IV 2) we see that the vectors Φ and Ψ must vary in two totally orthogonal subspaces.

Case A non-constant.

We may suppose that $n - 1$ directions (u^α) are positive and the last one (v) negative. We may take $A = v$ since A depends only on v .

$$\partial_a P = v \cdot \partial_a M, \quad \partial_v P = -v \cdot \partial_v M.$$

The integration of the first gives $P = vM + V(v)$. Then the last one may be written $2v\partial_v M + M + V'(v) = 0$. The integration of this linear differential equation of the first order gives

$$M = 1/\sqrt{v}(M_1(u^a) - \int (V'(v)dv/2\sqrt{v}))$$

$$P = \sqrt{v}(M_1(u^a) - \int (V'(v)dv/2\sqrt{v})) + V(v).$$

But one must have $\partial_a M \cdot \partial_v M = 0$, or $\partial_a M_1 \cdot (M + V') = 0$, or

$$(1) \quad M_1 \cdot \partial_a M_1 + \partial_a M_1 \cdot (V'\sqrt{v} - \int (V'dv/2\sqrt{v})) = 0.$$

If we write $V_1 = V'\sqrt{v} - \int (V'dv/2\sqrt{v})$, we get $\partial_a M_1 \cdot \partial_v V_1 = 0$.

Then, by IV 2), $M_1 = M_0 + M_2$, $V_1 = V_0 + V_2$, M_0 and V_0 being constant vectors, $M_2(u^a)$ and $V_2(v)$ vectors in two totally orthogonal subspaces \mathfrak{M} and \mathfrak{B} . (1) may be written then $\partial_a M_2 \cdot (M_0 + V_0 + M_2) = 0$. One may suppose $M_0 = 0$ (it may be absorbed by V) and $V_0 \in \mathfrak{M}$. Then $V_0 + M_2$ has a constant length. Let $U(u^a)$ denote a unit vector of \mathfrak{M} , $M = k \cdot U(u^a) - V_0$, k being a constant.

Let us consider $V_0 + V_2 = V'\sqrt{v} - \int (V'dv/\sqrt{v})$. The projection on \mathfrak{B} , since V_2 may vary, does not impose any conditions on the components of V' in \mathfrak{B} . But, by projection on \mathfrak{M} , the vector W' , the projection of V' on \mathfrak{M} , must satisfy $W'\sqrt{v} - \int W'dv/\sqrt{v} = \text{const.}$ Hence W' is a constant vector W_0 . Then $V' = V'_3 + W_0$ ($V'_3 \in \mathfrak{B}$, $W_0 \in \mathfrak{M}$). In order to avoid the integrations, we take $1/\sqrt{v} = w$ as a new variable, and for V a derivative W' ($W \in \mathfrak{B}$). M may be absorbed by W . Writing $-M_0$ instead of V_0 , we get the following parametric representations:

$$M = w(kU(u^a) + M_0) + w(W - W'w)$$

$$P = (1/w)(kU(u^a) + M_0) + (1/w)(W - W'w)$$

where $U(u^a)$ is a unit vector in \mathfrak{M} , $M_0 \in \mathfrak{M}$, k is a numerical constant, and $W \in \mathfrak{B}$ subspace orthogonal to \mathfrak{M} .

These formulae are a generalization of the formulae for the correspondence between revolution surfaces.

Conformal hyperspherical representation of a hypersurface.

If the manifold (M) is a hypersurface, it can be mapped by parallel tangent planes on the hypersphere of radius 1. Then the "operator" is the

second fundamental tensor a^{α}_{β} , the principal directions are the tangents to the lines of curvature, and the characteristic roots a_s are the principal curvatures.

If this correspondence is conformal, all the principal curvatures must be equal in absolute value. We have here an example of decomposition of properties. In the classical case of a surface in the 3-space the condition of conformal correspondence with its spherical image is that the surface be minimal. But the reason for this fact is not "the sum of the principal curvatures is 0," but "the principal curvatures have the same absolute value." These two properties, identical in the 3-space, become different in higher spaces. Here is one of the principal interests of the generalizations; to show the real reason for the properties, and to separate what is natural from what is due to coincidence.

For a hyperspherical representation the principal directions are real. Since the hypersphere is not a translation manifold, the case " A constant" gives only the trivial homothety. In the case " A non constant" $U(u^a)$ must run over all the unit vectors of an $(N-1)$ -dimensional subspace, for convenience the subspace spanned by the x_j axes, $j \geq 2$. Then in the parametric representation, $M_0 = 0$, $w = \sum_{j \geq 2} x_j^2$ (k gives only a homothety, and we take $k = 1$). The other coordinate $x_1 = w(W - W'w)$ (W being here a scalar function of w) must satisfy $x_1^2 + w^2 = 1$, or $W'w - W = \sqrt{1 - w^2}/w$. Taking $1/w = \cosh \phi$, the solution is: $W = -(\sinh 2\phi - 2\phi + a)/4 \cosh \phi$. The locus of P is therefore a *revolution manifold*, whose meridian has the following parametric representation.

$$r = \sum_{j \geq 2} x_j^2 = 1/w = \cosh \phi, \quad x_1 = (W'w + W) \cdot 1/w = \phi - (a/2).$$

This meridian is the *catenary* $r = \cosh(x_1 + a/2)$, and the manifold is the natural generalization of the minimal surface of revolution.

We have here a second example of a loss of freedom by passage from ordinary surfaces to higher manifolds: the surfaces with conformal, non-homothetic spherical representations are all the minimal surfaces, which depend on an arbitrary analytic function. For the hypersurfaces one obtains essentially one solution.

We may compare this fact to the conformal mappings of an euclidean space itself. For the plane the solutions depend on an arbitrary analytic function; but for the n -space ($n \geq 3$) the solution is given by the anallagmatic group which depends on a finite number of parameters, $(n+1)(n+2)/2$.³

³ Klein, F., "Vorlesungen über höhere Geometrie," p. 197, Berlin, 1926.

VII. Three-Dimensional Case.

There still remains the case where one principal direction is real (we may take it positive without loss of generality) and the two others are isotropic.

Holonomic case.

Calling u, v, w the parameters one must have

$$\partial_u P = A \partial_u M, \quad \partial_v P = A \cdot e^{i\theta} \partial_v M, \quad \partial_w P = A \cdot e^{-i\theta} \partial_w M$$

and

$$ds^2 = E \cdot du^2 + 2H \cdot dv \cdot dw.$$

The compatibility conditions are

$$A(1 - e^{-i\theta}) \partial_{uw} M = \partial_u (A e^{-i\theta}) \partial_w M - \partial_w A \partial_u M.$$

$$A(1 - e^{i\theta}) \partial_{uv} M = \partial_u (A e^{i\theta}) \partial_v M - \partial_v A \cdot \partial_u M.$$

$$2Ai \cdot \sin \theta \cdot \partial_{vw} M = \partial_v (A e^{-i\theta}) \partial_w M - \partial_w (A e^{i\theta}) \partial_v M.$$

By scalar multiplication with $\partial_u M$, $\partial_v M$, $\partial_w M$, and since

$$\partial_u M \cdot \partial_v M = \partial_u M \cdot \partial_w M = (\partial_v M)^2 = (\partial_w M)^2 = 0, \quad (\partial_u M)^2 = E,$$

$\partial_v M \cdot \partial_w M = H \neq 0$, we obtain two identities and

$$\partial_v (A e^{-i\theta}) = \partial_w (A e^{i\theta}) = \partial_u (A e^{i\theta}) = \partial_u (A e^{-i\theta}) = \partial_u H = 0$$

$$\partial_u E/E = -2(\partial_w A/A(1 - e^{-i\theta})); \quad \partial_v E/E = -2(\partial_v A/A(1 - e^{i\theta})).$$

Changing the variables v and w , we can take $A e^{-i\theta} = w^2$, $A e^{i\theta} = v^2$. Then $A = vw$, $e^{i\theta} = v/w$, $e^{-i\theta} = w/v$; and the integration of the two conditions for E gives $E = \Psi_1(u, v)(w - v/w)^2 = \Psi_2(u, w)((w - v)/v)^2$. Comparing them; $E = \phi(u)(v - w)^2/v^2 w^2$. Changing the variable u we can take $E = (v - w)^2/v^2 w^2$, $H = H(v, w)$. Then the compatibility conditions take the simpler form

$$\partial_{uw} M = (v/w(w - v)) \partial_u M, \quad \partial_{uv} M = (w/v(v - w)) \partial_u M, \quad \partial_{vw} M = 0$$

and are solved by $M = ((w - v)/vw)U(u) + V(v) + W(w)$.

It remains to study the conditions for ds^2_M ; for E it is sufficient to take $U'^2 = 1$. The conditions $\partial_u M \cdot \partial_v M = \partial_u M \cdot \partial_w M = 0$ become

$$-(1/v^2)UU' + U'V' = + (1/w^2)UU' + U'W' = 0.$$

If

$$U^2 = \Psi(u), \quad UU' = \frac{1}{2}\Psi'(u),$$

and

$$(U'/\Psi'(u)) \cdot (2v^2V') = 1, \quad (U'/\Psi'(u)) (2w^2W') = -1.$$

If

$$U'/\Psi'(u) = U_1(u), \quad 2v^2V' = V_1(v), \quad 2w^2W' = W_1(w),$$

we obtain by derivation

$$U'_1V_1 = 0, \quad U_1V'_1 = 0, \quad U'_1W_1 = 0, \quad U_1W'_1 = 0.$$

If \mathfrak{B} is the linear subspace spanned by the vectors U_1 , \mathfrak{B}_y the totally orthogonal subspace,

$$V' \in \mathfrak{B}_y, \quad W'_1 \in \mathfrak{B}_1, \quad V_1 = V_1^y + V_1^0, \quad W_1 = W_1^y + W_1^0,$$

$$U_1 \in \mathfrak{B}_x, \quad V_1^0 = \text{const} \in \mathfrak{B}_x, \quad W_1^0 = \text{const} \in \mathfrak{B}_y, \quad V_1^y \in \mathfrak{B}_y, \quad W_1^y \in \mathfrak{B}_y.$$

We must have also $V_1^0U'_1 = W_1^0U'_1 = 0$. If \mathfrak{B}_{x_1} is the plane spanned by the two constant vectors V_1^0 , W_1^0 and \mathfrak{B}_{x_2} the totally orthogonal subspace in \mathfrak{B}_x , $U'_1 \in \mathfrak{B}_{x_2}$. We obtain therefore the following decomposition of U, V, W :

$$\begin{array}{lll} \text{in } \mathfrak{B}_{x_1}: & U_1^0 & V_1^0 & W_1^0 \\ \text{in } \mathfrak{B}_{x_2}: & \bar{U}_1 & 0 & 0 \\ \text{in } \mathfrak{B}_y: & 0 & V_1^y & W_1^y \end{array}$$

with $U_1^0 \cdot V_1^0 = -U_1^0 \cdot W_1^0 = 1$.

We omit the integration constant for V and W (it is equivalent to a translation), and obtain the following decompositions of U, V, W :

$$\begin{array}{lll} \text{in } \mathfrak{B}_{x_1}: & U_1^0\Psi(u) + U & -V_1^0/2v & -W_1^0/2w \\ \text{in } \mathfrak{B}_{x_2}: & \bar{U} & 0 & 0 \\ \text{in } \mathfrak{B}_y: & U_y^0 & V_y & W_y. \end{array}$$

The isotropy condition $(\partial_v M)^2 = 0$ is: $(U^2/v^2) - V'^2 - 2(U \cdot V'/v^2) = 0$ which gives a quadratic equation in $\Psi(u)$, with coefficients depending only on v . Then $\Psi(u)$ is an absolute constant a , and U^1 may be absorbed in U_1^0 . The situation is the same in \mathfrak{B}_y and \mathfrak{B}_{x_1} , and, with an obvious change of notation, we obtain the following decompositions of U, V, W :

$$\begin{array}{lll} \text{in } \mathfrak{B}_x: & U_x & 0 & 0 \\ \text{in } \mathfrak{B}_y: & U_y^0 & V_y & W_y. \end{array}$$

With the additional scalar conditions:

$$\begin{aligned} U^2 &= U_x^2 + (U_y^0)^2 = a^2 & U'^2 &= 1 \\ V_y'^2 - (2/v^2) V_y' \cdot U_y^0 + a^2/v^4 &= 0 \\ W_y'^2 - (2/w^2) W_y' \cdot U_y^0 + a^2/v^4 &= 0 \end{aligned}$$

all our conditions are satisfied, and the parametric representations of the two 3-manifolds are

$$M = (w - v/vw)U + V + W;$$

$$P = (w - v) \cdot U + \int v^2 V' dv + \int w^2 W' dw.$$

In order to obtain real manifolds we must have A and θ real. Since $v = \sqrt{A} e^{i\theta/2}$, $w = \sqrt{A} e^{-i\theta/2}$, v and w must be complex conjugate for a real point. Hence we must take for V and W the same analytic vector function \mathfrak{F} , having complex conjugate values for two complex conjugate values of the variable (i. e. the Maclaurin's expansion of $\mathfrak{F}(z)$ has all its coefficients real), and for $U(u)$ a purely imaginary vector $iU_r(u)$. Then

$$M = 2 \sin (\theta/2) / A \cdot U_r(u) + \mathfrak{F}(\sqrt{A} e^{-i\theta/2}) + \mathfrak{F}(\sqrt{A} e^{i\theta/2}).$$

Non-holonomic case.

We may suppose that $ds_M^2 = (du^2)^2 + 2(du^i)(du^{i*})$. Let us recall that

$$\Gamma_{\rho\sigma}^\tau = \{\rho^\tau{}_\sigma\} - \frac{1}{2}(\Omega_{\rho,\sigma}^\tau + \Omega_{\sigma,\rho}^\tau + \Omega_{\rho\sigma}^\tau), \quad \Omega_{\rho\sigma}^\tau = \Omega_\rho{}^\tau{}_\sigma.$$

Then the conditions II a) and II b) may be written:

$$\Omega^{i*}{}_{ia} = \Omega^{i*}{}_{i^*a} = 0$$

$$\Omega^{i*}{}_{ai^*} + \Omega^i{}_{ai} = i \tan \theta/2 \cdot \Omega^a{}_{ii^*}$$

$$\partial_i(\log A) = (1 - e^{i\theta})\Omega^a{}_{ai}$$

$$\partial_{i^*}(\log A) = (1 - e^{i\theta})\Omega^a{}_{ai^*}$$

$$\partial_{i^*}(Ae^{i\theta}) = \partial_{i^*}(Ae^{-i\theta}) = 0.$$

$$\partial_a(\log(Ae^{i\theta})) = e^{i\theta} \cdot \tan \theta/2 \cdot \Omega^a{}_{ii^*}$$

$$\partial_a(\log(Ae^{-i\theta})) = e^{-i\theta} \cdot \tan \theta/2 \cdot \Omega^a{}_{i^*i}$$

The conditions $\Omega^{i*}{}_{ia} = \Omega^{i*}{}_{i^*a} = 0$ suggest computing non-symmetrically with respect to i and i^* . The system $\{\partial_a f = 0, \partial_i f = 0\}$ is complete; we call w a first integral of it, and u_1 and v_1 two independent first integrals of the equation $\partial_{i^*} f = 0$. Now $\partial_a = a\partial_{u_1} + b\partial_{v_1}$, $\partial_i = a'\partial_{u_1} + b'\partial_{v_1}$, $\partial_{i^*} = c\partial_w$. But, since $\Omega^{i*}{}_{ai^*} = 0$, $\partial_{i^*a} - \partial_{ai^*}$ must be a linear combination of ∂_a and ∂_{i^*} . It is $(a \cdot \partial_{u_1} c + b \cdot \partial_{v_1} c)\partial_w - c(\partial_w a \cdot \partial_{u_1} + \partial_{u_1} b \cdot \partial_{v_1})$. One must have $\partial_w a/a = \partial_w b/b$, or $a = b\phi(u_1, v_1)$. Now $\partial_a = b(\partial_{v_1} + \phi(u_1, v_1)\partial_{u_1})$. Let v be a first integral of $\partial_{v_1} f + \phi(u_1, v_1)\partial_{u_1} f = 0$, and u a function of u_1 and v_1 independent of v . Then $\partial_a = a_1\partial_u$, $\partial_i = b_1\partial_u + c_1\partial_v$, $\partial_{i^*} = c\partial_w$. The conditions $\partial_a M \cdot \partial_i M$

$= \partial_a M \cdot \partial_i M = (\partial_i M)^2 = (\partial_i M)^2 = 0$ give $(\partial_w M)^2 = \partial_u M \cdot \partial_w M = 0$. Thus, for the manifold (M) , $ds^2 = Edu^2 + 2Fdu dv + Gdv^2 + 2Hdv dw$, with $b_1/c_1 = -F/E$, $G = F/E^2$. $ds^2 = (Edu + Fdv)^2/E + 2Hdv dw$, and $\partial_i = c(\partial_v - (F/E)\partial_u)$. The fundamental system becomes, in the holonomic parameters u, v, w :

$$\partial_u P = A \cdot \partial_u M.$$

$$\partial_v P = A \cdot (F/E) \cdot (1 - e^{i\theta}) \cdot \partial_u M + A \cdot e^{i\theta} \cdot \partial_r M.$$

$$\partial_w P = A \cdot e^{-i\theta} \cdot \partial_w M.$$

$\partial_i(Ae^{i\theta}) = 0$ gives $\partial_w(Ae^{i\theta}) = 0$, or $A = e^{-i\theta} \cdot a(u, v)$. $\partial_i(Ae^{-i\theta}) = 0$ gives $E \cdot \partial_v(Ae^{-i\theta}) - F \cdot \partial_u(Ae^{-i\theta}) = 0$, or $E \cdot \partial_v(ae^{-2i\theta}) - F \cdot \partial_u(ae^{-2i\theta}) = 0$. (2)

Let us write the compatibility conditions:

- a) $(u, w) A(1 - e^{-i\theta})\partial_{uw}M = \partial_u(Ae^{-i\theta})\partial_w M - \partial_w A\partial_u M$.
- b) $(v, w) (AF/E)(1 - e^{i\theta})\partial_{vw}M + A(e^{i\theta} - e^{-i\theta})\partial_{vw}M$
 $+ \partial_w((AF/E)(1 - e^{-i\theta}))\partial_u M - \partial_v(Ae^{-i\theta})\partial_w M = 0$
- c) $(u, v) (AF/E)(1 - e^{i\theta})\partial_u^2 M + A(e^{i\theta} - 1)\partial_{uv}M + Q\partial_u M + \partial_u(Ae^{i\theta})\partial_v M$
 $= 0$

where

$$Q = \partial_u((AF/E)(1 - e^{i\theta})) - \partial_v A.$$

By scalar multiplication with $\partial_u M$ (1), $\partial_v M$ (2), $\partial_w M$ (3), we obtain:

- a) 1. $A(1 - e^{-i\theta})\partial_w E = -2 \cdot \partial_w A \cdot E$.
2. $A(1 - e^{-i\theta})(\partial_w F + \partial_u H) = 2H\partial_u(Ae^{-i\theta}) - 2F\partial_w A$.
3. already verified.
- b) 1. $(AF/2E)(1 - e^{i\theta})\partial_w E + E\partial_w((AF/E)(1 - e^{i\theta}))$
 $+ (A/2)(e^{i\theta} - e^{-i\theta})(\partial_w F - \partial_u H) = 0$
2. $(AF/2E)(1 - e^{i\theta})(\partial_w F + \partial_u H) + F\partial_w((AF/E)(1 - e^{i\theta}))$
 $+ (A/2)(e^{i\theta} - e^{-i\theta})\partial_w(F^2/E) - H\partial_v(Ae^{-i\theta}) = 0$
3. already verified.
- c) 1. $(AF/2E)(1 - e^{i\theta})\partial_u E + (A/2)(-1 + e^{i\theta})\partial_v E + QE$
 $+ (F^2/E)\partial_u(Ae^{i\theta}) = 0$.

$$2. (AF/2E) (1 - e^{i\theta}) (2\partial_u F - \partial_v E) + \frac{1}{2}A(e^{i\theta} - 1)\partial_u(F^2/E) \\ + QF + (F^2/E)\partial_u(Ae^{i\theta}) = 0.$$

$$3. (AF/2E) (1 - e^{i\theta})\partial_w E + A/2(1 - e^{i\theta})(\partial_u H - \partial_w F) \\ - H\partial_u(Ae^{i\theta}) = 0.$$

1) Taking $A = a(u, v)e^{-i\theta}$ and integrating a) 1 with respect to w , we obtain: $\sqrt{E} = b(u, v) \cdot (1 - e^{i\theta})$.

2) Using these values of A and E , b) 1 and c) 3 may be written

$$(e^{i\theta} + 1)\partial_u H = (1 - e^{i\theta})\partial_w F + 2iFe^{i\theta}\partial_w \theta = R$$

and

$$(e^{i\theta} + 1)H\partial_u a/a = -[(1 - e^{i\theta})\partial_w F + 2iFe^{i\theta}\partial_w \theta] = -R.$$

If $\partial_u a = 0$, then $R = 0$,

$$F = c(u, v)(1 - e^{i\theta})^2 \text{ and } Edu + Fdv = D(u, v, w)du_1$$

u_1 being a suitable function of u and v . Replacing u by u_1 , we get $F = 0$, and we are actually in the holonomic case. If $a \neq 0$, we may write

$$a\partial_u H + H \cdot \partial_u a = 0, \text{ or } H = \phi(v, w)/a(u, v).$$

3) Substituting in a) 2 the values of H and $\partial_u H$ obtained above, and factoring out R , we obtain ($R \neq 0$ unless we are in the holonomic case)

$$e^{-2i\theta} \cdot \partial_u a - \partial_u(ae^{-2i\theta}) = 0.$$

Therefore, $\partial_u \theta = 0$, $\theta = \theta(v, w)$.

4) c) 2 is the same as c) 1 multiplied by F/E .

5) Since $\partial_u \theta = 0$, (2) can be written $\partial_v(ae^{-2i\theta}) - F/E \cdot e^{2i\theta} \cdot \partial_u a = 0$. And c) 1 becomes $\partial_u(aFe^{-i\theta}/b(1 - e^{i\theta})^2) = \partial_v(abe^{-i\theta})$ with the solution $abe^{-i\theta} = \partial_u \Phi$, $aFe^{-i\theta}/b(1 - e^{i\theta})^2 = \partial_v \Phi$. Therefore $\partial_u(\Phi \cdot e^{i\theta})$ does not depend on w and may be written $\partial_u c(u, v)$. Hence

$$\Phi = c(u, v)e^{-i\theta} + \lambda_1(v, w), \quad F/E = \partial_v \Phi / abe^{i\theta} = \partial_v \Phi / \partial_u \Phi$$

(since $ab = \partial_u c$, and $E = b^2(1 - e^{i\theta})^2$). Since $F/E = \partial_v(ae^{-2i\theta})/\partial_u(ae^{-2i\theta})$, the jacobian $D(\Phi, ae^{-2i\theta})/D(u, v) = 0$, and $\Phi = \Phi(ae^{-2i\theta}, w)$ (3). But $\Phi = c(u, v)e^{-i\theta} + \lambda_1(v, w)$ (4). Φ must depend on u (unless $ab = 0$ which is impossible). From its form (3) it depends on u through $ae^{-2i\theta}$, and from its form (4) only through the first term. Hence $c(u, v)e^{-i\theta} = \Psi(ae^{-2i\theta})$. If θ does not depend on w , neither does F/E , and we are in the holonomic

case. If θ depends on w , we must take $c = \sqrt{a}$, $\Psi = \sqrt{ae^{-2i\theta}}$, and $\Phi = \sqrt{ae^{2i\theta}} + W(w)$. Then $b = \partial_u a / (2a\sqrt{a})$. We may now suppose that $a = u$. This change of variable does not change the forms of the ds^2 and of the fundamental system. Then $b = 1/(2u\sqrt{u})$, $E = 1/4u^3(1 - e^{i\theta})$, $F/E = -2iu\partial_v\theta$, $F = i/2u^2(1 - e^{i\theta})\partial_v\theta$.

6) Substituting these values in c) 3 we obtain

$$Hu = (i(1 - e^{i\theta})^3/2(1 + e^{i\theta}))\partial_v\theta$$

which is compatible with the results of 2).

7) With these values b) 2 is identically verified.

Thus all the projections of the compatibility conditions on the tangent 3-plane are verified; four of our nine conditions were consequences of the five others since we had used four conditions in the "study with α , i , i^* ."

With $\lambda(v, w) = e^{i\theta}$, the fundamental system is:

$$\partial_u P = u \cdot \lambda \cdot \partial_u M.$$

$$\partial_v P = u \cdot \partial_v M + 2u^2(1 - 1/\lambda) \cdot \partial_v \lambda \cdot \partial_u M.$$

$$\partial_w P = u \cdot \lambda^2 \cdot \partial_w M.$$

The compatibility conditions are now

$$u\lambda(\lambda - 1)\partial_{uw}M = u \cdot \lambda'_w \cdot \partial_u M - \lambda^2 \cdot \partial_w M.$$

$$u(1 - \lambda^2) \cdot \partial_{uv}M = 2u^2(1 - 1/\lambda)\lambda'_v \cdot \partial_{uv}M = 2u \cdot \lambda \cdot \lambda'_v \cdot \partial_w M \\ - 2u^2 \cdot \partial_w((1 - 1/\lambda)\lambda'_v) \cdot \partial_u M.$$

$$u(\lambda - 1) \cdot \partial_{vw}M - 2u^2(1 - 1/\lambda) \cdot \lambda'_v \cdot \partial_w M = \partial_v M + u \cdot \lambda'_v(3 - 4/\lambda)\partial_u M.$$

By scalar multiplication with the normal vectors n_a , we obtain the following necessary and sufficient conditions for the coefficients of the second forms

$$a_{Muw,a} = a_{Mvw,a} = 0, \quad \lambda a_{Muv,a} = 2u \cdot \lambda'_v \cdot a_{Muw,a}.$$

And the ds^2 must be

$$ds^2_M = \frac{(1 - \lambda)^2}{u} \left[\left(\frac{du}{2u} - \frac{\lambda'_v}{\lambda} dv \right)^2 + \frac{1 - \lambda}{1 + \lambda} \left(\frac{\lambda''_{vw}}{\lambda} - \frac{\lambda'_v \lambda'_w}{\lambda^2} \right) dv dw \right].$$

Using θ the necessary and sufficient conditions are

$$ds^2_M = \frac{(1 - e^{i\theta})^2}{u} \left[\left(\frac{du}{2u} - i\theta'_v dv \right)^2 + i \cdot \tan \frac{\theta}{2} \cdot e^{-i\theta} \cdot \theta''_{vw} \cdot dv dw \right].$$

$$a_{Muw,a} = a_{Mvw,a} = 0, \quad a_{Muv,a} = 2iu \cdot \theta'_v \cdot a_{Muw,a}.$$

VIII. General Holonomic Case.

We consider the fundamental system

$$\partial_a P = A \cdot \partial_a M, \quad \partial_\lambda P = -A \cdot \partial_\lambda M, \quad \partial_k P = A e^{i\theta_k} \cdot \partial_k M, \quad \partial_{k^*} P = A e^{-i\theta_k} \partial_{k^*} M$$

with $ds^2_M = g_{aa}(du^a)^2 + \dots + g_{\lambda\lambda}(du^\lambda)^2 + \dots + 2g_{kk^*} du^k \cdot du^{k^*} + \dots$

the coordinate directions being holonomic.

Compatibility conditions.

a) In k, k^* : $A i \cdot \sin \theta_k \cdot \partial_{kk^*} M = \partial_k (A e^{-i\theta_k}) \cdot \partial_{k^*} M - \partial_{k^*} (A e^{i\theta_k}) \cdot \partial_k M.$

By scalar multiplication with the tangent vectors we obtain:

$$\partial_a g_{kk^*} = \partial_\lambda g_{kk^*} = \partial_j g_{kk^*} = \partial_k (A e^{-i\theta_k}) = \partial_{k^*} (A e^{i\theta_k}) = 0.$$

b) In j, k : $A (e^{i\theta_j} - e^{i\theta_k}) \cdot \partial_j M = \partial_j (A e^{i\theta_k}) \cdot \partial_k M - \partial_k (A e^{i\theta_j}) \cdot \partial_j M.$

By $\partial_a M, \partial_\lambda M, \partial_\rho M, \partial_j M$: identically verified.

By $\partial_{j^*} M$: $A/2 (e^{i\theta_j} - e^{i\theta_k}) \cdot \partial_k g_{jj^*} = -\partial_k (A e^{i\theta_j}).$

But, because of a), $\partial_k (A e^{i\theta_j}) = 0$, and similarly $\partial_k (A e^{-i\theta_j}) = 0.$

c) In α, k : $A (1 - e^{i\theta_k}) \cdot \partial_\alpha M = \partial_\alpha M - \partial_\alpha (A e^{i\theta_k}) \cdot \partial_k M.$

By $\partial_\beta M, \partial_\lambda M, \partial_j M, \partial_k M$: identically verified.

By $\partial_a M$: $\frac{1}{2} (\partial_k g_{aa}/g_{aa}) = \partial_k A/A (1 - e^{i\theta_k}).$

By $\partial_{k^*} M$: $\partial_a (A e^{i\theta_k}) = 0$ (since $\partial_a g_{kk^*} = 0$).

d) We obtain only, in λ, k :

$$\frac{1}{2} (\partial_k g_{\lambda\lambda}/g_{\lambda\lambda}) = \partial_k A/A (1 - e^{i\theta_k}) \text{ and } \partial_\lambda (A e^{i\theta_k}) = 0.$$

e) In α, β : $\partial_\alpha A \cdot \partial_\beta M - \partial_\beta A \cdot \partial_\alpha M = 0.$

Therefore $\partial_\alpha A = \partial_\beta A = 0$, if there exists at least two positive directions. In the same way $\partial_\lambda A = \partial_{\mu} A = 0$, if there exists at least two negative directions.

f) In α, λ : $2A \cdot \partial_\alpha M = \partial_\alpha A \cdot \partial_\lambda M - \partial_\lambda A \cdot \partial_\alpha M.$

We see that $A e^{i\theta_k}$ depends only on u^k , $A e^{-i\theta_k}$ only on u^{k^*} . If they are not constant we can write $A e^{i\theta_k} = (u^k)^2$, $A e^{-i\theta_k} = (u^{k^*})^2$. Then $A = u^k u^{k^*}$. And since it is impossible to have also $A = u^j u^{j^*}$, $A e^{i\theta_j}$ and $A e^{-i\theta_j}$ must be constant. Then A is constant which is contrary to $A = u^k u^{k^*}$. Hence there is only one couple of isotropic coordinate directions.

Case with one couple of isotropic principal directions.

Let $u^k = v$, $u^{k*} = w$. Then $A = vw$, and $\partial_a A = \partial_\lambda A = 0$ are verified.
 $e^{i\theta} = v/w$, $e^{-i\theta} = w/v$.

Consider the compatibility conditions

$$\partial_{v\lambda} M = 0, \quad \partial_{va} M = -(w/v(w-v))\partial_a M, \quad \partial_{wa} M = (v/w(w-v)) \cdot \partial_a M,$$

and the similar ones with λ instead of α . Their integration gives

$$M = ((w-v)/vw)U(u^a, \dots, u^\lambda, \dots) + V(v) + W(w).$$

The ds^2 -conditions impose that:

1) U satisfies the same conditions as M in VI (real principal directions). Only the two first cases can happen; V case 3) cannot happen since $\partial_a A = \partial_\lambda A = 0$.

2) There exist two supplementary subspaces \mathfrak{B}_x and \mathfrak{B}_y with the following decompositions of U, V, W (see: VII, holonomic case):

$$\text{In } \mathfrak{B}_x: \quad U_x \qquad 0 \qquad 0.$$

$$\text{In } \mathfrak{B}_y: \quad U_y^0 \quad (\text{const.}) \quad V \quad W$$

3) $U^2 = a^2 = \text{constant}$,

$$V'^2 - (2/v^2)V' \cdot \dot{U}_y^0 + (a^2/v^4) = 0, \quad W'^2 - (2/r^2) \cdot W' \cdot U_y^0 + (a^2/w^4) = 0.$$

Several couples of isotropic principal directions.

Then all the angles of rotation θ_k are constant, and, since $\partial_\kappa(Ae^{i\theta_j}) = 0$, A depends only on the parameters corresponding to the real principal directions. But $\partial_a(Ae^{i\theta_k}) = \partial_\lambda(Ae^{i\theta_k}) = 0$. Therefore A is also a constant. After a homothety we can take $A = 1$, and the manifolds are *isometric*.

The compatibility conditions are $\partial_{a\lambda} M = 0$, $\partial_{aj} M = \partial_{\lambda j} M = 0$ (because $\theta_j \neq k\pi$: in the case $\theta_j = k\pi$, these directions are real). And $\partial_{jj^*} M = 0$, $\partial_{jk} M = 0$ (if $\theta_j \neq \theta_k$).

Denoting by j_1, j_2, \dots the indices corresponding to the same angle we get the following parametric representation:

$$M = M_1(u^a, \dots) + M_2(u^\lambda, \dots) + M_j(u^{j_1}, \dots, u^{j_n}) + M^*_{j_1}(u^{j^*_1}, \dots, u^{j^*_n}) \\ + \dots + M_k(u^{k_1}, \dots, u^{k_p}) + M^*_{k_1}(u^{k^*_1}, \dots, u^{k^*_p})$$

with the following conditions due to the ds^2 . We do not consider the conditions $g_{j_1 j^*_2} = 0$ which are not necessary (see III, 4).

a) $M_1, M_2, M_j, M^*_{j_1}, \dots, M_k$ and $M^*_{k_1}$ vary in mutually orthogonal linear subspaces.

b) The manifolds (M_j) and $(M^*_{j_1}), \dots, (M_k)$ and $(M^*_{k_1})$, are totally isotropic, and complex conjugate in the real case.

IX. Isometric Manifolds.

We shall now study the case where the manifolds (M) and (P) are not only in conformal, but in isometric correspondence.

Preliminary formula: principal normal to a curve on a manifold.

We consider the manifold $M(u^a, \dots)$ in the euclidean N -space, and on it the curve C , $u^a = u^a(s)$, s being the arc on C . The primes denote the derivations with respect to s : $g_{a\beta} \cdot u'^a \cdot u'^\beta = 1$ by definition of the arc. The tangent unit vector of C is $\vec{t} = dM/ds = \partial_a M \cdot u'^a$. If \vec{n} denotes the principal normal unit vector, and R the first radius of curvature of C , we have by definition $d^2M/ds^2 = \vec{n}/R$, and, by the Frenet formulae of the manifold

$$\vec{n}/R = d/ds(\partial_a M \cdot u'^a) = \partial_a M \cdot u''^a + \partial_{a\beta} M \cdot u'^a \cdot u'^\beta \\ = (u''^\lambda + \Gamma^\lambda_{a\beta} u'^a \cdot u'^\beta) \cdot \partial_\lambda M + a^i_{a\beta} \cdot u'^a \cdot u'^\beta \cdot \vec{n}_i.$$

If \vec{n} is tangent to the manifold, $a^i_{a\beta} \cdot u'^a \cdot u'^\beta = 0$ (asymptotic curves). If \vec{n} is normal, $u''^\lambda + \Gamma^\lambda_{a\beta} \cdot u'^a \cdot u'^\beta = 0$ (geodesics).

A property of the geodesics.

Since here the two manifolds are isometric, the geodesics are corresponding curves. We consider two corresponding geodesics, with the same arc s , and with tangent unit vectors $\vec{t}(s)$, $\vec{t}'(s)$, with angle V between them; $\cos V = \vec{t} \cdot \vec{t}'$.

$\sin V \cdot dV/ds = \vec{t} \cdot \vec{n}'/R' + \vec{t}' \cdot \vec{n}/R = 0$, since \vec{n} and \vec{n}' are normal vectors. Hence the angle of rotation, which is applied to a tangent vector t on (M) in order to obtain the corresponding tangent vector on (P) , is stationary for a motion of M in the direction of \vec{t} .

If the manifolds are surfaces, and if the correspondence between tangent planes is a proper rotation, the angle of rotation is the same in all directions, and is evidently constant on every part of the surface whose two points can be joined by geodesics.

Constancy of the rotation angles.

This fact may be generalized. We shall prove that, when the rotation is canonically decomposed into rotations in mutually orthogonal planes, with angles θ_k , these angles are constant. We denote by b^α_λ the linear operator of rotation, by p^α the contravariant components of a tangent vector. By rotation it is transformed to $q^\alpha = b^\alpha_\lambda \cdot p^\lambda$. The angle V between these vectors satisfies $\cos V = g_{\alpha\beta} p^\beta q^\alpha = g_{\alpha\beta} b^\alpha_\lambda p^\lambda p^\beta$. By parallel displacement along the geodesic tangent to p^α at M , the covariant derivative $p^\mu \nabla_\mu p^\alpha$ is zero. The covariant derivative of the scalar $\cos V$ is the ordinary derivative which is also zero as proved before. Hence $g_{\alpha\beta} p^\beta p^\lambda \nabla_\mu (b^\alpha_\lambda) = 0$. This relation is true for every vector p^α ; being a polynomial in the p^α 's, its coefficients must be all zero; $g_{\alpha\beta} \nabla_\mu (b^\alpha_\lambda) = 0$. (The parenthesis around the indices β, μ, λ denotes, as usual, that one has to take the sum of the 6 terms corresponding to all the permutations of the 3 indices β, μ, λ .)

If, at the point M , we take geodesic coordinates with the principal directions as coordinate directions, the first covariant derivatives are, at M , equal to the ordinary derivatives. Hence, using indices $a, b, c \dots$, in order not to interfere with the conventions of III, $g_{ab} \nabla_a (b^a_c) = 0$ can be written

$$g_{ab} \partial_a s_a + g_{ba} \partial_a s_b + g_{da} \partial_b s_d + g_{ad} \partial_b s_a + g_{bd} \partial_a s_b + g_{db} \partial_a s_d = 0$$

or

$$g_{ab} \partial_a (s_a + s_b) + g_{ad} \partial_b (s_a + s_d) + g_{bd} \partial_a (s_b + s_d) = 0. \quad (E)$$

But here ${}_a s = 1$, ${}_b s = -1$, ${}_k s = e^{i\theta_k}$, ${}_{k^*} s = e^{-i\theta_k}$, $g_{aa} = g_{\lambda\lambda} = g_{kk^*} = 1$ the other g_{ab} being zero. With two Greek indices, or with two indices j, k , (E) is verified. With k, k^* , and any other index a : $g_{kk^*} \partial_a (e^{i\theta_k} + e^{-i\theta_k}) = 0$, $\partial_a \theta_k = 0$. With k, k, k^* : $g_{kk^*} \partial_k (e^{i\theta_k} + e^{-i\theta_k}) + g_{kk^*} \cdot \partial_k (e^{i\theta_k} + e^{-i\theta_k}) = 0$.

Hence $\partial_k \theta_k = 0$, and similarly $\partial_{k^*} \theta_k = 0$. Therefore the θ_k are constant. This result is also true if some θ 's are equal.

We shall now show that, in the 3- and 4-dimensional cases, the principal directions are holonomic, so that we are concerned with a case already studied in VIII.

We shall take, as in the 3-dimensional non-holonomic case,

$$ds^2_M = (du^1)^2 + \dots + (du^{\lambda})^2 + \dots + 2(du^s)(du^{s^*}) + \dots,$$

and write the compatibility conditions in terms of the Ω 's. We shall not use the constancy of the θ_k (the computation is only slightly simplified by using this fact) but prove it a second time.

3-dimensional case.

Putting $A = 1$ in the conditions (1) of VII, non-holonomic case, we get

$$\Omega^a_{ai} = \Omega^a_{ai^*} = \Omega^{i^*}_{ia} = \Omega^{i^*}_{i^*a} = \partial_i \theta = \partial_{i^*} \theta = 0.$$

$$\partial_a (\log (A e^{i\theta})) = i \partial_a \theta = e^{i\theta} \tan (\theta/2) \Omega^a_{i^*i}.$$

$$\partial_a (\log (A e^{i\theta})) = -i \partial_a \theta = e^{-i\theta} \tan (\theta/2) \Omega^a_{i^*i}.$$

Since $\theta \neq \kappa\pi$, $\partial_a \theta = 0$, $\Omega^a_{i^*i} = 0$. Hence θ is a constant, and the conditions for the Ω 's prove immediately that there exists 3 parameters u, v, w such that

$$\partial_a = a_1 \partial_u, \quad \partial_i = a_2 \partial_v, \quad \partial_{i^*} = a_3 \partial_w.$$

Therefore the principal directions are holonomic.

Compatibility conditions in the general case.

We do not write the intermediary computations which are long but without mathematical difficulties.

1) In α, β .

α, β, γ : verified

$$\lambda : \Omega^\lambda_{\alpha\beta} = 0$$

$$k : (e^{i\theta_k} - 1) \Omega^k_{\alpha\beta} = 0. \quad \text{Hence } \Omega^k_{\alpha\beta} = 0.$$

2) In λ, μ :

$$\alpha : \Omega^\alpha_{\lambda\mu} = 0.$$

λ, μ, ν : verified.

$$k : \Omega^k_{\lambda\mu} = 0.$$

3) In α, λ :

$$\alpha : \Omega^{(a)}_{(a)\lambda} = 0.$$

$$\beta : \Omega^\beta_{a\lambda} + \Omega^{a\beta}_{\beta\lambda} = 0 \text{ (using 1), } \lambda).$$

$$\lambda : \Omega^{(\lambda)}_{a(\lambda)} = 0.$$

$$\mu : \Omega^\mu_{a\lambda} + \Omega^\lambda_{a\mu} = 0 \text{ (using 2), } \alpha).$$

$$k : e^{i\theta_k} \Omega^k_{a\lambda} + \Omega^{\lambda}_{k^*a} + \Omega^a_{k^*\lambda} = 0.$$

$$k^* : e^{-i\theta_k} \Omega^{k^*}_{a\lambda} + \Omega^\lambda_{ka} + \Omega^a_{k\lambda} = 0.$$

4) In α, k (or λ, k).

$$\alpha : \Omega^{(a)}_{(a)k} = 0, \text{ and } \Omega^{(\lambda)}_{(\lambda)k} = 0.$$

$$\beta : \Omega^\beta_{ak} + \Omega^{a\beta}_{\beta k} = 0, \text{ and } \Omega^\mu_{\lambda k} + \Omega^\lambda_{\mu k} = 0.$$

$$\lambda : \Omega^\lambda_{ka} = 0, \text{ and } \Omega^a_{k\lambda} = 0.$$

Hence, using 3), $k^* : \Omega^{k^*}_{a\lambda} = \Omega^k_{a\lambda} = 0$.

$$k : \Omega^{a_{k^*}}_{k^*k} + \Omega^{(k^*)}_{(k^*)a} + \Omega^{(k)}_{(k)a} = 2(\partial_a(e^{i\theta_k})/e^{i\theta_k} - 1).$$

$$\text{Similarly: } \Omega^{a_{kk^*}}_{kk^*} + \Omega^{(k)}_{(k)a} + \Omega^{(k^*)}_{(k^*)a} = -2(\partial_a(e^{-i\theta_k})/e^{-i\theta_k} - 1)$$

and the same with λ , but with $e^{i\theta_k} + 1$ and $e^{-i\theta_k} - 1$ in denominators.

$$j : (2e^{i\theta_j} - e^{i\theta_k} - 1)\Omega^j_{ak} = (e^{i\theta_k} - 1)(\Omega^{a_{k^*}}_{k^*k} + \Omega^k_{j^*a})$$

$$\text{and } (2e^{i\theta_j} - e^{i\theta_k} + 1)\Omega^j_{\lambda k} = (e^{i\theta_k} + 1)(\Omega^{\lambda_{j^*}}_{j^*k} + \Omega^k_{j\lambda}).$$

5) In k, k^* .

$$\alpha : \Omega^{a_{kk^*}}_{kk^*}(1 - \cos \theta_k) = i \sin \theta_k (\Omega^{(k)}_{(k)a} + \Omega^{(k^*)}_{(k^*)a}).$$

Comparing with 4), $k : \Omega^{a_{kk^*}}_{kk^*} = \Omega^{(k)}_{(k)a} = \Omega^{(k^*)}_{(k^*)a} = \partial_a \theta_k = 0$

$$\lambda : \text{in the same way } \Omega^{\lambda_{kk^*}}_{kk^*} = \Omega^{(k)}_{(k)\lambda} = \Omega^{(k^*)}_{(k^*)\lambda} = \partial_\lambda \theta_k = 0.$$

$$k : \partial_k \theta_k = 0 \text{ (the terms in } \Omega \text{ disappear).}$$

Similarly $\partial_k \theta_k = 0$.

$$j : \Omega_{kk^*}^j (e^{i\theta_j} - \cos \theta_k) + (\Omega_{j^*(k)}^{(k)} + \Omega_{j^*(k^*)}^{(k^*)}) i \sin \theta_k = 0.$$

$$j^* : \Omega_{kk^*}^{j^*} (e^{-i\theta_j} - \cos \theta_k) + (\Omega_{j(k)}^{(k)} + \Omega_{j(k^*)}^{(k^*)}) i \sin \theta_k = 0.$$

6) In k, j :

$$\alpha : \Omega_{kj}^\alpha (2 - e^{i\theta_k} - e^{i\theta_j}) = (e^{i\theta_j} - e^{i\theta_k}) (\Omega_{\alpha k}^{j^*} + \Omega_{\alpha j}^{k^*}).$$

$$\lambda : \Omega_{kj}^\lambda (-2 - e^{i\theta_k} - e^{i\theta_j}) = (e^{i\theta_j} - e^{i\theta_k}) (\Omega_{\lambda k}^{j^*} + \Omega_{\lambda j}^{k^*}).$$

$$k^* : (e^{i\theta_j} - e^{-i\theta_k}) \Omega_{kj}^{k^*} = 0.$$

$$k : \Omega_{(k)j}^{(k)} + \Omega_{(k^*)j}^{k^*} + \Omega_{kk^*}^{j^*} = 2\partial_j (e^{-i\theta_k}) / e^{i\theta_j} - e^{i\theta_k}.$$

Exchanging k and k^* ,

$$\Omega_{(k)j}^{(k)} + \Omega_{(k^*)j}^{(k^*)} + \Omega_{kk^*}^{j^*} = 2\partial_j (e^{-i\theta_k}) / e^{i\theta_j} - e^{i\theta_k}.$$

Comparing with 5), j ,

$$\partial_j \theta_k = 0, \quad \Omega_{kk^*}^j = \Omega_{(k)j}^{(k)} + \Omega_{(k^*)j}^{(k^*)} = 0$$

(even if $\theta_j = -\theta_k$).

$$1 : \Omega_{jk}^1 (2e^{i\theta_j} - e^{i\theta_k} - e^{i\theta_j}) = (e^{i\theta_j} - e^{i\theta_k}) (\Omega_{1k}^{j^*} + \Omega_{1j}^{k^*}).$$

We see that *the angles θ_k are constant*.

Holonomy in the 4-dimensional case.

We shall now prove that, in the 4-dimensional case, we have enough symbols Ω equal to zero in order to make the principal directions holonomic. In the general case, when the four characteristic roots are distinct, we notice that every system of two equations extracted from the system

$$\{\partial_{\alpha} f = \partial_{\lambda} f = \partial_k f = \partial_{k^*} f = 0\} \quad (\text{or } \{\partial_k f = \partial_{k^*} f = \partial_j f = \partial_{j^*} f = 0\})$$

is complete.

LEMMA. *If every system of two equations extracted from*

$$\{\partial_{\alpha} f = \partial_{\lambda} f = \partial_k f = \partial_{k^*} f = 0\},$$

is complete, then the non-holonomic derivatives $\partial_{\alpha}, \partial_{\lambda}, \partial_k, \partial_{k^}$ are proportional to holonomic derivatives.*

The system $\{\partial_\lambda f = \partial_k f = \partial_{k^*} f = 0\}$, being complete, has a first integral u . Let us denote by v, w, t three independent first integrals of $\partial_a f = 0$. Then

$$\partial_a = a\partial_u, \quad \partial_\lambda = a_1\partial_v + b_1\partial_w + c_1\partial_t, \quad \partial_k = a_2\partial_v + b_2\partial_w + c_2\partial_t,$$

$$\partial_{k^*} = a_3\partial_v + b_3\partial_w + c_3\partial_t.$$

$$\partial_{a\lambda} - \partial_{\lambda a} = a(\partial_u a_1 \cdot \partial_v + \partial_u b_1 \cdot \partial_w + \partial_u c_1 \cdot \partial_t) - (a_1 \cdot \partial_v a + b_1 \cdot \partial_w a + c_1 \partial_t a \partial_u).$$

But, since $\Omega^k_{a\lambda} = \Omega^{k^*}_{a\lambda} = 0$, $\partial_{a\lambda} - \partial_{\lambda a}$ is a linear combination of ∂_a and ∂_λ . Hence we must have $\partial_u a_1/a_1 = \partial_u b_1/b_1 = \partial_u c_1/c_1$, whence

$$a_1 = \phi_1(u, v, w, t) a'_1(v, w, t), \quad b_1 = \phi_1(u, v, w, t) b'_1(v, w, t), \quad c_1 = \phi_1(u, v, w, t) c'_1(v, w, t)$$

and similarly for $a_2, b_2, c_2, a_3, b_3, c_3$. Now

$$\begin{aligned} \partial_\lambda &= \phi_1 \cdot (a'_1 \partial_v + b'_1 \partial_w + c'_1 \partial_t), & \partial_k &= \phi_2 \cdot (a'_2 \partial_v + b'_2 \partial_w + c'_2 \partial_t), \\ \partial_{k^*} &= \phi_3 \cdot (a'_3 \partial_v + b'_3 \partial_w + c'_3 \partial_t). \end{aligned}$$

Hence the equations $\{\partial_\lambda f = \partial_k f = \partial_{k^*} f = 0\}$ do not contain the variable u . We make the same reasoning for them; we isolate ∂_λ , and, since $\Omega^\lambda_{kk^*} = 0$, we change the variables v, w, t into v', w', t' such that

$$\begin{aligned} \partial_\lambda &= \phi_1 \cdot l \cdot \partial_{v'}, & \partial_k &= \phi_2 \cdot (l_1(v', w', t') \partial_{w'} + m_1(v', w', t') \partial_{t'}), \\ \partial_{k^*} &= \phi_3 \cdot (l_2(v', w', t') \partial_{w'} + m_2(v', w', t') \partial_{t'}). \end{aligned}$$

Since

$$\Omega^{k^*}_{k\lambda} = \Omega^{k^*}_{k^*j} = 0, \quad \partial_{v'}(\phi_2 l_1)/\phi_2 l_1 = \partial_{v'}(\phi_2 m_1)/\phi_2 m_1$$

and

$$\partial_{t'}(\phi_3 l_2)/\phi_3 l_2 = \partial_{t'}(\phi_3 m_2)/\phi_3 m_2$$

or $\partial_{v'} l_1/l_1 = \partial_{v'} m_1/m_1$ and $\partial_{t'} l_2/l_2 = \partial_{t'} m_2/m_2$, which are solved by

$$\begin{aligned} l_1 &= \psi_1(v', w', t') \cdot l'_1(w', t'), & m_1 &= \psi_1 \cdot m'_1(w', t') \\ l_2 &= \psi_2(v', w', t') \cdot l'_2(w', t'), & m_2 &= \psi_2 \cdot m'_2(w', t'). \end{aligned}$$

Now $\partial_k = \phi_2 \cdot \psi_1 \cdot (l'_1(w', t') \partial_{w'} + m'_1 \cdot \partial_{t'})$, $\partial_{k^*} = \phi_3 \cdot \psi_2 \cdot (l'_2 \cdot \partial_{w'} + m'_2 \cdot \partial_{t'})$. Changing the variables w' and t' we can write

$$\begin{aligned} l'_1 \cdot \partial_{w'} + m'_1 \cdot \partial_{t'} &= n_1(w', t') \partial_{w''}, \\ l'_2 \cdot \partial_{w'} + m'_2 \cdot \partial_{t'} &= n_2(w', t') \partial_{t''}. \end{aligned}$$

Hence, in the 4-dimensional case (and more generally, in any dimension, in the case where there are only two couples of isotropic principal directions with different rotation angles, at most one positive and at most one negative principal direction) we are in the holonomic case.

If there are several principal directions corresponding to the same characteristic root, some conditions $\Omega = 0$ do not hold; but, by compensation, we have more freedom for choosing the principal directions, every direction of the linear space spanned by the principal directions corresponding to the same characteristic root being principal.

Therefore: *there exist principal directions which form a holonomic coordinate system, when there are at most two different rotation angles.*

Remark. This result is probably true in general. The conditions of the type 6), 1, with all the possible combinations of the 6 indices j, j^*, k, k^*, l, l^* , form a system of $6(6-1)/2 = 30$ linear homogeneous equations in 30 unknowns Ω . Their determinant is probably $\neq 0$ (if $\theta_j, -\theta_j, \theta_k, -\theta_k, \theta_l, -\theta_l$ are all different; otherwise it is already proved). Hence $\Omega'_{jk} = 0, \dots$, and the holonomy is proved by the lemma.

PRINCETON UNIVERSITY.

SUL CALCOLO DELLE SOLUZIONI DEI PROBLEMI AL CONTORNO PER LE EQUAZIONI LINEARI DEL SECONDO ORDINE DI TIPO ELLITTICO.*

By LUIGI AMERIO.

Introduzione.

Nei suoi *Appunti di Analisi Superiore* il Picone¹ tratta i classici problemi al contorno per le equazioni differenziali lineari a derivate parziali seguendo un unico criterio, basato su una nuova e suggestiva interpretazione, da Lui stesso indicata, della formula di Green.

In virtù di tale criterio, la soluzione dei problemi considerati si deduce immediatamente dalla preventiva risoluzione di alcuni problemi di Analisi che il Picone pone in modo ben preciso calendosi, se si considerano le equazioni lineari del secondo ordine, di tipo ellittico, delle seguenti considerazioni.

Sia

$$(1) \quad E(u) = \sum a_{ik} (\partial^2 u / \partial x_i \partial x_k) + \sum b_i (\partial u / \partial x_i) + cu = f \quad (a_{ik} = a_{ki})$$

un'equazione di tal natura, in m variabili (x_1, \dots, x_m) , e si suppongano le a_{ik} , b_i funzioni reali di classe 2 e 1 rispettivamente,² le c , f reali e continue in un dominio limitato τ di S_m . Inoltre il determinante $\|a_{ik}\|$ sia unitario, ciò che manifestamente non lede la generalità.

* Lavoro eseguito presso l'Istituto Nazionale per le Applicazioni del Calcolo. Received September 21, 1946.

¹ M. Picone, *Appunti di Analisi Superiore*, Rondinella, Napoli, 1940, pp. 752-65.

² Seguendo G. Ascoli, (*Equazioni alle derivate parziali dei tipi ellittico e parabolico*, Sansoni, Firenze, 1935, p. 53) diremo che una funzione u è di classe k in un dominio τ se è continua insieme alle sue derivate parziali di ordine $\leq k$ in ogni punto interno a τ e se è possibile prolungare la u e le sue derivate parziali di ordine $\leq k$ anche sul contorno σ di τ in modo da risultare continue in tutto τ . Diremo inoltre che la u è di classe kH in τ se le sue derivate parziali di ordine $\leq k$ soddisfano in tutto τ e una condizione di Hölder. Scriveremo H in luogo di oH .

Una varietà σ a $m-1$ dimensioni si dirà di classe k in un suo punto M se risulta definita, in un intorno completo di M da una rappresentazione (invertibile) delle coordinate dei suoi punti come funzioni di classe k di $m-1$ parametri variabili in un certo dominio, con matrice funzionale non nulla. Si dirà poi di classe k una varietà σ la quale sia di classe k in ogni suo punto.

Introdotta l'espressione differenziale aggiunta di $E(u)$:

$$(2) \quad E^*(w) = \Sigma (\partial^2 (a_{ik}w) / \partial x_i \partial x_k) - \Sigma (\partial (b_i w) / \partial x_i) + cw \\ = \Sigma a_{ik} (\partial^2 w / \partial x_i \partial x_k) + \Sigma b_i^* (\partial w / \partial x_i) + c^* w,$$

supponiamo che il contorno σ di τ soddisfi a quelle condizioni che assicurano la validità della formula di Green per ogni coppia u, w di funzioni di classe 2 in τ :

$$- \int_{\tau} \{u E^*(w) - w E(u)\} d\tau = \int_{\sigma} \{u (\partial w / \partial \nu) - w (\partial u / \partial \nu) - Luw\} d\sigma,$$

dove ν è la direzione conormale,³ orientata verso l'interno di τ , e si è posto

$$L = \Sigma \cos (nx_i) \{b_i - \Sigma (\partial a_{ik} / \partial x_k)\},$$

n essendo la normale a σ , orientata anch'essa verso l'interno di τ .

Risulta allora, se u è un integrale della (1) di classe 2 in τ ,

$$(3) \quad - \int_{\tau} u E^*(w) d\tau = \int_{\sigma} \{u ((\partial w / \partial \nu) - Lw) - (\partial u / \partial \nu) w\} d\sigma - \int_{\tau} f w d\tau$$

e da questa classica relazione il Picone deduce le seguenti proposizioni, che riferirò, per semplicità, al problema di Dirichlet.

a) *Se si conosce una successione $\{v_r\}$ di integrali dell'equazione aggiunta*

$$(4) \quad E^*(v) = 0$$

la quale sia chiusa,⁴ su σ , della (4) si ricava il sistema di Fischer-Riesz:

$$(5) \quad \int_{\sigma} \{u ((\partial v_r / \partial \nu) - L v_r) - (\partial u / \partial \nu) v_r\} d\sigma - \int_{\tau} f v_r d\tau = 0$$

e quindi si conoscono i coefficienti di Fourier dell'incognita $\partial u / \partial \nu$ rispetto alla successione $\{v_r\}$. La funzione $\partial u / \partial \nu$ resta perciò determinata nei punti di σ .

b) *Se si conosce una successione $\{w_r\}$ di funzioni di classe 2 in τ tali che la successione $\{E^*(w_r)\}$ sia chiusa, in τ , dalla (3) si ricava il sistema di Fischer-Riesz:*

³ ν è la retta di parametri direttori $\nu_i = \Sigma a_{ik} \cos (ux_k)$. Osserviamo che $\partial u / \partial \nu$ risulta *proporzionale* alla derivata della funzione u rispetto alla direzione ν : benchè impropriamente chiameremo, seguendo l'uso comune, tale funzione la *derivata conormale* di u .

⁴ In questo enunciato, e nei seguenti, basterà considerare la chiusura rispetto alla totalità delle funzioni a quadrato sommabile nell'assegnato insieme di definizione. Noi però considereremo, in seguito, la chiusura anche rispetto alla totalità delle funzioni sommabili.

$$(6) \quad - \int_{\tau} u E^*(w_r) d\tau = \int_{\sigma} \{u((\partial w_r / \partial \nu) - L w_r) - (\partial u / \partial \nu w_r)\} d\sigma - \int_{\tau} f w_r d\tau$$

e, quindi, note, su σ , le u e $\partial u / \partial \nu$, resta determinata la u nei punti di τ .

Osserviamo che, mediante le successioni $\{v_r\}$, $\{w_r\}$ considerate in a), b), si ricava prima la $\partial u / \partial \nu$ su σ e successivamente la u in τ . È possibile però enunciare una terza proposizione relativa alla contemporanea determinazione di tali funzioni.

Cominciamo per questo, col ricordare la seguente estensione, indicata dal Picone,⁵ della classica teoria degli sviluppi in serie di funzioni ortogonali e normali.

Siano f_1, \dots, f_p , p funzioni reali definite rispettivamente nelle varietà I_2, \dots, I_p di S_m , aventi ciascuna un proprio numero $\leq m$ di dimensioni. Consideriamo tali funzioni come componenti di un vettore \mathbf{F} dello spazio S_p . Se ω è un altro vettore, di componenti $\omega_1, \dots, \omega_p$ in I_1, \dots, I_p rispettivamente, chiameremo *prodotto integrale* di \mathbf{F} per ω , e lo indicheremo col simbolo

$$(\mathbf{F}, \omega)$$

la somma

$$(\mathbf{F}, \omega) = \sum_1^p \int_{I_h} f_h \omega_h dI_h.$$

Il prodotto integrale (\mathbf{F}, \mathbf{F}) , si dirà la *norma integrale* di \mathbf{F} e una successione $\{\mathbf{z}_r\}$ di vettori si dirà *ortogonale* e *normale* se risulterà

$$(\mathbf{z}_r, \mathbf{z}_s) = \begin{cases} 1 & \text{per } r = s \\ 0 & \text{per } r \neq s. \end{cases}$$

Da una qualsiasi successione $\{\omega_r\}$ di vettori di norma integrale finita si può sempre, mediante l'estensione di un noto procedimento, dedurne una equivalente, $\{\mathbf{z}_r\}$, ortogonale e normale.

Consideriamo ora un sistema di equazioni di Fischer-Riesz:

$$(7) \quad (\mathbf{F}, \omega_r) = c_r$$

dove \mathbf{F} è un vettore incognito e le c_r sono costanti assegnate, che chiameremo i *coefficienti di Fourier* di \mathbf{F} rispetto alla successione $\{\omega_r\}$. Sostituito al sistema (7) l'equivalente sistema

$$(8) \quad (\mathbf{F}, \mathbf{z}_r) = d_r,$$

⁵ M. Picone, *loc. cit.*¹, pp. 632-34, Vedasi anche, dello stesso Autore, "Vedute unitarie sul calcolo delle soluzioni delle equazioni alle derivate parziali della fisica matematica," *Atti I Convegno di Matematica Applicata*, Roma, 1936.

(dove le d_r risultano note, perchè ciascuna è combinazione lineare di un numero finito di c_r) vale il teorema di Fischer-Riesz: *condizione necessaria e sufficiente perchè il sistema (7) ammetta una soluzione F di norma integrale finita è che converga la serie Σd^2_r* . In tal caso la serie di vettori $\Sigma d_r z_r$ converge in media a un vettore F' per il quale $(F', \omega_r) = c_r$, cioè, indicata con $z_{r,h}$ la componente di z_r in I_h , la serie $\Sigma d_r z_{r,h}$ converge in media nella varietà I_h alla componente f'_h del vettore F' . Indicato poi con F'' il più generale vettore, di norma integrale finita, ortogonale a tutti i vettori ω_r , cioè tale che risulti

$$(9) \quad (F'', \omega_r) = 0,$$

la soluzione F del sistema (7) (nella totalità dei vettori di norma integrale finita) è data dalla somma

$$(10) \quad F = F' + F''.$$

Il vettore F'' dipende linearmente da un certo numero di costanti arbitrarie. Inoltre, per una qualsiasi soluzione F del sistema (7) vale la disuguaglianza di Bessel: $\Sigma d^2_r \leq (F, F)$. Infine condizione necessaria e sufficiente perchè la successione $\{\omega_r\}$ sia *chiusa* rispetto alla totalità dei vettori di norma integrale finita, cioè il vettore F'' sia quasi ovunque nullo, è che per un qualsiasi vettore F di norma integrale finita valga l'eguaglianza di Parseval: $\Sigma d^2_r = (F, F)$.

Osserviamo anche che, se le componenti $\omega_{r,h}$ di ω_r sono limitate, nel sistema (7) si possono supporre le incognite f_h sommabili nelle rispettive varietà I_h . Vale anche in tal caso la (10), dove F' è una particolare soluzione, di componenti sommabili, del sistema (7) e F'' il più generale vettore, di componenti sommabili, ortogonale alla successione $\{\omega_r\}$. La possibilità di ottenere il vettore F' come somma della serie $\Sigma d_r z_r$, calcolata con un conveniente metodo di sommazione, richiede però un particolare studio, analogamente a quanto avviene per le serie di Fourier.

Ciò premesso, considerando sempre il problema di Dirichlet, possiamo enunciare la seguente proposizione:

c) Sia $\{w_z\}$ una successione di funzioni di classe 2 in τ tale che la successione di vettori $\{\omega_r\}$, di componenti $-E^*(w_r)$ in τ , w_r su σ , risulti chiusa. Detto G il vettore di componenti u in τ , $\partial u / \partial \nu$ su σ , si ricava dalla (6) il sistema di Fischer-Riesz:

$$(11) \quad (G, \omega_r) = \int_{\sigma} u((\partial w_r / \partial \nu) - L w_r) d\sigma - \int_{\tau} f w_r d\tau.$$

Sono perciò noti i coefficienti di Fourier dell'incognito vettore \mathbf{G} rispetto alla successione $\{\omega_r\}$ è quindi tale vettore risulta individuato.

Negli enunciati a), c) si suppone manifestamente che la soluzione esista e sia unica. Osserviamo infatti che, se manca il teorema di unicità non esistono successioni $\{v_r\}$, $\{w_r\}$ soddisfacenti alle condizioni di chiusura poste in a), c).

Indicheremo perciò le condizioni cui debbono soddisfare le successioni $\{v_r\}$, $\{w_r\}$ per poter trattare, attraverso i sistemi (5), (11), anche questi casi.

d) Le successioni $\{v_r\}$, $\{w_r\}$ debbono esser tali che, se la soluzione del problema considerato non esiste, i sistemi (5), (11) non ammettano soluzioni. Se manca il teorema di unicità, le derivate conormali delle autosoluzioni debbono esser caratterizzate dall'essere ortogonali, su σ , alla successione $\{v_r\}$, le autosoluzioni dall'essere i corrispondenti vettori \mathbf{G} ortogonali alla successione $\{\omega_r\}$.

Proposizioni analoghe alle a), b), c), d), si possono enunciare per gli altri problemi al contorno relativi alla (1).

Si presenta ora la questione di costruire effettivamente le successioni $\{v_r\}$, $\{w_r\}$ dotate delle proprietà richieste in a), b), c), d), di risolvere cioè i problemi porti dal Picone.

A tale proposito, il Picone,⁶ considerando le sole proposizioni a), b) ha dimostrato, come conseguenza della nota chiusura in τ della successione di funzioni $x_1^{\alpha_1} \cdots x_m^{\alpha_m}$, con $\alpha_i = 0, 1, \dots$, rispetto alla totalità delle funzioni sommabili in τ , che, se la (1) è a coefficienti costanti e se (p'_1, \dots, p'_m) non è soluzione dell'equazione

$$(1?) \quad \sum a_{ik} p_i p_k - \sum b_i p_i + c = 0,$$

posto $w = e^{\sum p_i x_i}$ e inoltre

$$w_r = \left(\frac{\partial^{\alpha_1 + \dots + \alpha_m}}{\partial p_1^{\alpha_1} \cdots \partial p_m^{\alpha_m}} e^{\sum p_i x_i} \right)_{(p_i = p'_i)}, \quad (\alpha_i = 0, 1, \dots)$$

la successione $\{E^*(w_r)\}$ è chiusa rispetto alla totalità delle funzioni sommabili in τ .

⁶ Il Prof. Picone, dal quale gli eventi della guerra mi separarono, impedendomi per più di un anno di comunicare con Lui, mi partecipa ora di aver già Lui stesso assunto, per i problemi d'integrazione delle equazioni lineari a derivate parziali, di qualsivoglia ordine e tipo, il punto di vista al quale appartiene la proposizione c). Ciò formò oggetto di una sua conferenza—tenuta nel 1943 ma non pubblicata e della quale non ero al corrente durante la redazione del presente lavoro—avente per titolo: Questioni di topologia nella teoria dell'integrazione delle equazioni lineari a derivate parziali.

Quanto alla successione $\{v_r\}$, sempre supponendo la (1) a coefficienti costanti, dopo aver osservato che dalla funzione $v = e^{\sum p_i x_i}$, nell'ipotesi che i parametri p_i soddisfino alla (12), si può dedurre, in moda analogo a quanto si è fatto per la w (mediante le derivazioni di tutti gli ordini rispetto ai parametri p_i , effettuate tenendo conto della (12)) una successione $\{v_r\}$ di integrali della (4), ha proposto di studiare in general tale successione, rilevando che essa è sicuramente chiusa se la (1) è l'equazione di Laplace in 2 variabili e se τ è semplicemente connesso, perchè in tal caso si può prendere $v_r = (x_1 \pm x_2)^r$.

In un recente lavoro ⁷ ho risolto, nel caso generale, la questione posta dal Picone, dimostrando che la successione di funzioni $\{v_r\}$ da Lui indicata permette di risolvere i problemi al contorno per la (1), ma soltanto nell'ipotesi che il dominio τ abbia connessione ipersuperficiale semplice.

Nello stesso lavoro ho indicato un'altra successione $\{v_r\}$ che consente la risoluzione dei problemi considerati, qualunque sia la connessione di τ . Il procedimento di integrazione da me seguito si basa sulla conoscenza della soluzione fondamentale della (4), ben nota trattandosi di un'equazione a coefficienti costanti, e si può estendere a numerosi altri problemi al contorno relativi a equazioni lineari; inoltre in esso non si richiede la conoscenza della successione $\{w_r\}$ considerata nella proposizione b).

Nella prima parte di questo lavoro è appunto indicato come tale procedimento si applichi all'integrazione della (1), supponendola a coefficienti variabili.

Per questo cominciamo con l'ammettere che le funzioni a_{ik} , b_i , c si possano prolungare in un dominio τ' dotato di contorno σ' di classe 2, contenente τ nel suo interno, supponendo inoltre che in τ' le a_{ik} , b_i risultino di classe $2H$, $1H$ rispettivamente, la c di classe II .

Esiste allora ⁸ se R è un qualsiasi punto prefissato internamente a τ' e N un punto variabile in τ' , la soluzione fondamentale $F(N, R)$ dell'equazione $E^*(u) = 0$ (che possiamo scegliere in modo da soddisfare alla $E(u) = 0$ come funzione di R) e valgono, per ogni integrale della (1) di classe 2 in τ , le formule fondamentali di Green:

⁷ M. Picone, "Nuovi metodi risolutivi per i problemi d'integrazione delle equazioni lineari a derivate parziali e nuova applicazione della trasformata multipla di Laplace nel caso delle equazioni a coefficienti costanti." *Rendiconti Accademia delle Scienze di Torino*, 1940, pp. 413-26.

⁸ L. Amerio, "Sull'integrazione dell'equazione $\Delta_2 u - \lambda^2 u = f$ in un dominio di connessione qualsiasi," *Rendiconti Istituto Lombardo di Scienze e Lettere*, vol. 78 (1944-45).

$$(13) \quad k_m u(P) = \int_{\sigma} \{u(M) ((\partial F(M, P)/\partial v) - L(M)F(M, P)) \\ - (\partial u(M)/\partial v)F(M, P)\} d\sigma - \int_{\tau} f(M)F(M, P) d\tau$$

$$(14) \quad 0 = \int_{\sigma} \{u(M) ((\partial F(M, Q)/\partial v) - L(M)F(M, Q)) \\ - (\partial u(M)/\partial v)F(M, Q)\} d\sigma - \int_{\tau} f(M)F(M, Q) d\tau$$

dove k_m è una costante non nulla dipendente solo da m , P e Q indicano rispettivamente un punto interno e un punto esterno a τ , M è il punto, di τ o di σ , rispetto alle cui coordinate si effettuano la integrazioni.

Noi considereremo sempre i problemi al contorno per la (1) nella totalità degli integrali per cui valgano le (13), (14) intendendo, nel caso più generale, le funzioni $u(M)$, $\partial u(M)/\partial v$ definite quasi ovunque, su σ , come i limiti

$$\lim_{P \rightarrow M} u(P), \quad \lim_{P \rightarrow M} (\partial u(P)/\partial v),$$

dove $P \rightarrow M$ lungo la conormale v in M , e sommabili su σ .

Se poniamo, su σ , $u(M) = A(M)$, $\partial u(M)/\partial v = B(M)$, dalla (14) segue che A , B soddisfano all'equazione

$$(15) \quad 0 = \int_{\sigma} \{A(M) ((\partial F(M, Q)/\partial v) - L(M)F(M, Q)) \\ - B(M)F(M, Q)\} d\sigma - \int_{\tau} f(M)F(M, Q) d\tau$$

la quale appare perciò come *necessaria* per le coppie A , B di funzioni che coincidono rispettivamente con i valori assunti su σ da un integrale u della (1) e della sua derivata conormale $\partial u/\partial v$. Ma tale equazione è anche *sufficiente*: ho dimostrato infatti, facendo sulla natura di σ delle ipotesi largamente verificate nei casi pratici, che se A , B è una coppia di funzioni soddisfacenti alla (15), la funzione $u(P)$ definita internamente a τ dall'eguaglianza

$$(16) \quad k_m u(P) = \int_{\sigma} \{A(M) ((\partial F(M, P)/\partial v) - L(M)F(M, P)) \\ - B(M)F(M, P)\} d\sigma - \int_{\tau} f(M)F(M, P) d\tau$$

è un integrale della (1) soddisfacente su σ alle condizioni

$$u(M) = A(M), \quad \partial u(M)/\partial v = B(M).$$

Si può perciò dire che la (15) rappresenta l'equazione di compatibilità tra i valori assunti su σ da un integrale della (1) e dalla sua derivata conormale.

La (15) equivale poi a un sistema di Fischer-Riesz. Sia infatti $\{g_r(Q)\}$ una successione di funzioni continue, chiusa nel dominio $\bar{\tau} = \tau' - \tau + \sigma$ rispetto alla totalità delle funzioni sommabili in $\bar{\tau}$ e continue nell'interno di $\bar{\tau}$, totalità alla quale appartengono gli integrali a secondo membro della (15) (si può, ad esempio, prendere $g_r(Q) = x_1^{\alpha_1} \cdots x_m^{\alpha_m}$, con $\alpha_i = 0, 1, \dots$). La (15) equivale allora al sistema di infinite equazioni

$$\int_{\bar{\tau}} g_r(Q) \left[\int_{\sigma} \{A(M) ((\partial F(M, Q)/\partial v) - L(M)F(M, Q)) - B(M)F(M, Q)\} d\sigma - \int_{\tau} f(M)F(M, Q) d\tau \right] d\bar{\tau} = 0$$

cioè anche, posto

$$(17) \quad v_r(M) = \int_{\bar{\tau}} g_r(Q) F(M, Q) d\bar{\tau},$$

al sistema

$$(18) \quad \int_{\sigma} \{A(M) ((\partial v_r(M)/\partial v) - L(M)v_r(M)) - B(M)v_r(M)\} d\sigma - \int_{\tau} f(M)v_r(M) d\tau = 0.$$

Se si considera, ad esempio, il problema di Dirichlet e se vale, in tale problema, per la (1), il teorema di unicità si ha perciò che la successione $\{v_r\}$ data dalla (17) è chiusa su σ e soddisfa, essendo $v_r(M)$ un integrale della (4), alle condizioni richieste nella proposizione a). Se non vale il teorema di unicità, le autosoluzioni sono caratterizzate dall'essere le corrispondenti funzioni $B(M)$ ortogonali su σ alla successione $\{v_r\}$. Infine, se il problema non ammette soluzione, nemmeno il sistema (18) ammette soluzione. La successione $\{v_r\}$ soddisfa perciò alle condizioni poste in d).

La (16) dà poi l'integrale della (1) corrispondente a una determinata coppia di funzioni A, B .

Il procedimento di integrazione indicato non richiede l'uso della successione $\{w_r\}$ considerata nella proposizione b); in esso però si suppone esplicitamente nota la soluzione fondamentale $F(N, R)$ la cui effettiva costruzione è, in generale, assai ardua. Per ovviare a tale inconveniente ho indicato, nella seconda parte di questo lavoro un altro procedimento, in virtù del quale l'uso di tale funzione è completamente eliminato, essendo sufficiente di conoscere l'esistenza, ciò che avviene se i coefficienti a_{ik} , b_i , c soddisfano alle condizioni dianzi ricordate.

Per questo, poniamo nella (3) in luogo di w la funzione

$$w_r = x_1^{\alpha_1} \cdots x_m^{\alpha_m} \quad (\alpha_i = 0, 1, \cdots)$$

e, su σ , $u = A$, $\partial u / \partial \nu = B$. Otteniamo in tal modo il sistema di infinite equazioni di Fischer-Riesz:

$$(19) \quad - \int_{\tau} u E^*(w_r) d\tau = \int_{\sigma} \{A(\partial w_r / \partial \nu - L w_r) - B w_r\} d\sigma \\ - \int_{\tau} f w_r d\tau.$$

Anche tale sistema *caratterizza* gli integrali della (1). Ho dimostrato infatti che *se* u , A , B *sono tre funzioni per cui valgono le* (15), (16), *con* A , B *funzioni sommabili su* σ , *le funzioni* u , A , B *soddisfano al sistema* (19). *Viceversa, se* u , A , B *sono tre funzioni sommabili nei rispettivi insiemi di definizione e soddisfacenti al sistema* (19), *allora* A , B *soddisfano all'equazione* (15) *e* u , *ove se ne alteri al più il valore in un insieme di misura nulla, è data dalla* (16).

Per risolvere il problema di Dirichlet si procederà allora nel modo seguente. Si considerino le funzioni incognite u in τ , B in σ come componenti di un vettore \mathbf{G} , le funzioni $-E^*(w_r)$ in τ , w_r in σ come componenti di un vettore $\mathbf{\omega}_r$.

Si ottiene perciò che l'incognito vettore \mathbf{G} soddisfa al sistema (11) in cui si ponga A in luogo di u . Se vale poi il teorema di unicità la successione $\{\mathbf{\omega}_r\}$ così ottenuta risulta chiusa e soddisfa alle condizioni poste in c). Se tale teorema non vale, le autosoluzioni sono *caratterizzate* dall'essere i corrispondenti vettori \mathbf{G} ortogonali alla successione $\{\mathbf{\omega}_r\}$. Infine se manca la soluzione del problema di Dirichlet, anche il sistema (11) non ammette soluzione. La successione $\{\mathbf{\omega}_r\}$ soddisfa perciò alla condizioni poste in d).

Ragionamenti del tutto analoghi si fanno per gli altri problemi al contorno.

Osserviamo infine che, in luogo della successione $x_1^{\alpha_1} \cdots x_m^{\alpha_m}$ si può assumere, sotto condizioni assai larghe, come successione $\{w_r\}$ una qualsiasi successione chiusa in τ' rispetto alla totalità delle funzioni ivi sommabili. In ogni caso la corrispondente successione $\{E^*(w_r)\}$ risulta chiusa in τ rispetto alla totalità delle funzioni sommabili in τ e soddisfa perciò alla condizioni posta in b).

I. Primo procedimento di integrazione.

1. Ricordiamo che la parte principale $V(M, P)$ della soluzione fondamentale $F(M, P)$ è data da⁹

$$(20) \quad V(M, P) = \{Z(M, P)\}^{-(m-2/2)}, \quad \text{per } m > 2,$$

$$V(M, P) = -\frac{1}{2} \log Z(M, P), \quad \text{per } m = 2,$$

con

$$(21) \quad Z(M, P) = \Sigma A_{ik}(P) (x_i - \xi_i) (x_k - \xi_k)$$

dove $A_{ik}(P)$ è l'aggiunto di $a_{ik}(P)$, (x_1, \dots, x_m) e (ξ_1, \dots, ξ_m) sono le coordinate di M e P rispettivamente.

Ciò premesso, dimostriamo un lemma relativo al comportamento di alcune derivate della funzione $F(M, P)$.

Per questo, portata l'origine Ω delle coordinate di S_m in un punto M_0 di σ , cominciamo col supporre che, in M_0 , σ risulti di classe 2. Fatta coincidere la normale a σ in M_0 con l'asse x_m , in un conveniente intorno completo σ_0 di Ω . σ avrà perciò equazione $x_m = \phi(x_1, \dots, x_{m-1})$, con ϕ funzione di classe 2. Sia λ la conormale in Ω , ν la conormale in un generico punto M di σ_0 . Posto $a = \{\Sigma \nu^2_i(M_0)\}^{\frac{1}{2}}$, sia δ la distanza da Ω di un punto $P(\xi_1, \dots, \xi_m)$ interno a τ e situato sulla conormale in Ω . Siccome la conormale in Ω pentra in τ , il punto Q simmetrico di P rispetto a Ω sarà esterno a τ , se la distanza δ è sufficientemente piccola. Poniamo poi $\lambda = \delta/a$.

Dette allora $\delta' = \lambda'a$, $\delta'' = \lambda''a$ le ascisse di P e Q sulla conormale in Ω , sarà $\lambda' = \lambda$, $\lambda'' = -\lambda$ e quindi risulterà, in particolare,

$$\begin{aligned} \frac{\partial F(M, P)}{\partial \lambda'} &= \frac{\partial F(M, P)}{\partial \lambda}, \quad \frac{\partial F(M, Q)}{\partial \lambda''} = -\frac{\partial F(M, Q)}{\partial \lambda}, \\ \frac{\partial^2 F(M, Q)}{\partial \nu \partial \lambda'} &= \frac{\partial^2 F(M, Q)}{\partial \nu \partial \lambda''} = \frac{\partial^2 F(M, P)}{\partial \nu \partial \lambda} + \frac{\partial^2 F(M, Q)}{\partial \nu \partial \lambda}. \end{aligned}$$

LEMMA. Posto $s = \{\sum_1^{m-1} x_i^2\}^{\frac{1}{2}}$, valgono per le derivate della soluzione fondamentale le limitazioni

$$\begin{aligned} \frac{\partial F(M, P)}{\partial \nu} &= O\left(\frac{\lambda}{(s^2 + \lambda^2)^{m/2}} + \frac{1}{s^{m-2}}\right), \\ \frac{\partial F(M, P)}{\partial \lambda} &= O\left(\frac{\lambda}{(s^2 + \lambda^2)^{m/2}} + \frac{1}{s^{m-2}}\right), \\ \frac{\partial^2 F(M, P)}{\partial \nu \partial \lambda} + \frac{\partial^2 F(M, Q)}{\partial \nu \partial \lambda} &= O\left(\frac{\lambda}{(s^2 + \lambda^2)^{m/2}} + \frac{1}{s^{m-2}}\right) \end{aligned}$$

⁹ Vedasi, ad esempio, G. Ascoli, *loc. cit.*², pp. 70-76.

nelle quali, per $m = 2$, in luogo di $s^{-(m-2)}$, dovrà porsi $|\log s|$.

Dimostrazione. Supporremo $m > 2$, la dimostrazione per $m = 2$ essendo del tutto analoga.

a) Dimostriamo dapprima che per la funzione $V(M, P)$ data dalle (20), (21) valgono le limitazioni

$$(22) \quad \begin{aligned} \frac{\partial V(M, P)}{\partial \nu} &= O \left(\frac{\lambda}{(s^2 + \lambda^2)^{m/2}} + \frac{1}{s^{m-2}} \right), \\ \frac{\partial V(M, P)}{\partial \lambda} &= O \left(\frac{\lambda}{(s^2 + \lambda^2)^{m/2}} + \frac{1}{s^{m-2}} \right), \\ \frac{\partial^2 V(M, P)}{\partial \nu \partial \lambda} + \frac{\partial^2 V(M, Q)}{\partial \nu \partial \lambda} &= O \left(\frac{\lambda}{(s^2 + \lambda^2)^{m/2}} + \frac{1}{s^{m-2}} \right). \end{aligned}$$

Per questo, cominciamo col ricordare che risulta, per il parametro direttore ν_i della conormale ν ,

$$(23) \quad \nu_i = \Sigma a_{ik}(M) \cos (nx_k),$$

n essendo la normale in M , orientata verso l'interno di τ .

Si ha allora, per le (20), (21), ricordando che il determinante $\|a_{ik}\|$ è unitario,

$$(24) \quad \begin{aligned} \frac{\partial V(M, P)}{\partial \nu} &= -(m-2)Z(M, P)^{-m/2} \Sigma A_{ik}(P) (x_k - \xi_k) \nu_i \\ &= -(m-2)Z(M, P)^{-m/2} \Sigma A_{ik}(P) (x_k - \xi_k) a_{il}(M) \cos (nx_l) \\ &= -(m-2)Z(M, P)^{-m/2} \{ \Sigma (x_k - \xi_k) \cos (nx_k) \\ &\quad + \Sigma (a_{il}(M) - a_{il}(P)) A_{ik}(P) (x_k - \xi_k) \cos (nx_l) \} \end{aligned}$$

e, posto $\rho = \{ \sum_1^m x_i^2 \}^{\frac{1}{2}}$, $\cos \theta_i = x_i/\rho$, risulta

$$(25) \quad \begin{aligned} Z(M, P) &= \Sigma A_{ik}(P) (x_i - \xi_i) (x_k - \xi_k) \\ &= \Sigma A_{ik}(P) x_i x_k + \Sigma A_{ik}(P) \xi_i \xi_k - 2 \Sigma A_{ik}(P) \xi_i x_k \\ &= \rho^2 \Sigma A_{ik}(P) \cos \theta_i \cos \theta_k + \lambda^2 \Sigma A_{ik}(P) a_{im}(\Omega) a_{km}(\Omega) \\ &\quad - 2 \lambda \rho \Sigma A_{ik}(P) a_{im}(\Omega) \cos \theta_k, \end{aligned}$$

essendo, per la (23),

$$(26) \quad \xi_i = \lambda a_{im}(\Omega).$$

Si ha poi

(27) $\Sigma A_{ik}(P) a_{im}(\Omega) a_{km}(\Omega) = a_{mm}(\Omega) + \Sigma (A_{ik}(P) - A_{ik}(\Omega)) a_{im}(\Omega) a_{km}(\Omega)$
 ed è $a_{mm}(\Omega) \neq 0$ perchè la forma quadratica $\Sigma a_{ik} \xi_i \xi_k$ è definita positiva.

E' infine

$$(28) \quad \Sigma A_{ik}(P) a_{im}(\Omega) \cos \theta_k = \cos \theta_m + \Sigma (A_{ik}(P) - A_{ik}(\Omega)) a_{im}(\Omega) \cos \theta_k$$

e siccome risulta, indicando con α una costante positiva,

$$(29) \quad |x_m| = |\phi(x_1, \dots, x_{m-1})| \leq \alpha s^2,$$

$$(30) \quad \lim \rho/s = 1,$$

si ricava

$$(31) \quad |\cos \theta_m| \leq \beta s$$

con β costante positiva.

Dalle (25), (27), (28), (29), (30), (31), essendo la forma $\Sigma A_{ik} \xi_i \xi_k$ definita positiva, segue perciò che è possibile determinare $\delta > 0$ in modo che, per $|x_i| \leq \delta$, $0 < \lambda \leq \delta$, risulti

$$(32) \quad p(s^2 + \lambda^2) \leq Z(M, P) \leq q(s^2 + \lambda^2)$$

con $0 < p < q$.

Siccome poi si ha

$$(33) \quad |\cos(n x_k)| \leq \gamma s, \quad (k = 1, \dots, m-1)$$

con γ costante positiva, si ricava, per le (26), (29),

$$|\Sigma (x_k - \xi_k) \cos(n x_k)| \leq \eta(s^2 + \lambda),$$

dove η è una costante positiva, e quindi dalle (24), (32) segue la prima delle (22).

Consideriamo ora la derivata $\partial V(M, P)/\partial \lambda$. Si ha, per la (26), posto $A_{ikh}(P) = \partial A_{ik}(P)/\partial \xi_h$,

$$(34) \quad \begin{aligned} \frac{\partial V(M, P)}{\partial \lambda} = & -\frac{m-2}{2} Z(M, P)^{-m/2} \{ -2 \Sigma a_{hm}(\Omega) A_{hk}(P) (x_k - \xi_k) \\ & + \Sigma a_{hm}(\Omega) A_{ikh}(P) (x_k - \xi_k) (x_i - \xi_i) \} \\ = & (m-2) Z(M, P)^{-m/2} (x_m - \xi_m) \\ & - \frac{m-2}{2} Z(M, P)^{-m/2} \{ -2 \Sigma a_{hm}(\Omega) (A_{hk}(P) - A_{hk}(\Omega)) (x_k - \xi_k) \\ & + \Sigma a_{hm}(\Omega) A_{ikh}(P) (x_k - \xi_k) (x_i - \xi_i) \} \end{aligned}$$

e quindi dalle (26), (29), (32) segue la seconda delle (22).

Studiamo infine il comportamento della derivata $\partial^2 V(M, P)/\partial \nu \partial \lambda$.

Si ha, per la (26),

$$(35) \quad \frac{\partial^2 V(M, P)}{\partial \nu \partial \lambda} = \frac{m(m-2)}{2} Z(M, P)^{-(m+2/2)} \{ -2 \Sigma a_{hm}(\Omega) A_{hk}(P) (x_k - \xi_k) \\ + \Sigma a_{hm}(\Omega) A_{ikh}(P) (x_k - \xi_k) (x_i - \xi_i) \} \{ \Sigma a_{il}(M) A_{ik}(P) (x_k - \xi_k) \cos(nx_l) \} \\ - (m-2) Z(M, P)^{-m/2} \{ - \Sigma a_{hm}(\Omega) a_{il}(M) A_{ik}(P) \cos(nx_l) \\ + \Sigma a_{hm}(\Omega) a_{il}(M) A_{ikh}(P) (x_k - \xi_k) \cos(nx_l) \}$$

da cui, per le (24), (34),

$$(36) \quad \frac{\partial^2 V(M, P)}{\partial \nu \partial \lambda} = \frac{m(m-2)}{2} Z(M, P)^{-(m+2/2)} \{ -2(x_m - \xi_m) \\ - 2 \Sigma a_{hm}(\Omega) (A_{hk}(P) - A_{hk}(\Omega)) (x_k - \xi_k) \\ + \Sigma a_{hm}(\Omega) A_{ikh}(P) (x_k - \xi_k) (x_i - \xi_i) \} \{ \Sigma (x_k - \xi_k) \cos(nx_k) \\ + \Sigma a_{il}(M) (A_{ik}(P) - A_{ik}(M)) (x_k - \xi_k) \cos(nx_l) \} \\ - (m-2) Z(M, P)^{-m/2} \{ - \Sigma a_{im}(M) \cos(nx_l) \\ + \Sigma a_{hm}(\Omega) a_{il}(M) A_{ikh}(\Omega) (x_k - \xi_k) \cos(nx_l) \\ - \Sigma a_{hm}(\Omega) a_{il}(M) (A_{ih}(P) - A_{ih}(\Omega)) \cos(nx_l) \\ + \Sigma a_{hm}(\Omega) a_{il}(M) (A_{ikh}(P) - A_{ikh}(\Omega)) (x_k - \xi_k) \cos(nx_l) \} \\ = -m(m-2) Z(M, P)^{-(m+2/2)} \{ (x_m - \xi_m) (\Sigma (x_k - \xi_k) \cos(nx_k)) + \eta \} \\ - (m-2) Z(M, P)^{-m/2} \{ - \Sigma a_{im}(M) \cos(nx_l) \\ + \Sigma a_{hm}(\Omega) a_{il}(M) A_{ikh}(\Omega) (x_k - \xi_k) \cos(nx_l) \}$$

e risulta, per le (29), (32), (33),

$$(37) \quad \eta Z(M, P)^{-(m+2/2)} = O((s^2 + \lambda^2)^{-(m+2/2)} [(s^2 + \lambda)(s + \lambda)^2 \\ + (\lambda(s + \lambda) + (s + \lambda)^2)(s^2 + \lambda + (s + \lambda)^2)]) \\ = O\left(\frac{\lambda}{(s^2 + \lambda^2)^{m/2}} + \frac{1}{s^{m-2}}\right),$$

$$(38) \quad \xi Z(M, P)^{-m/2} = O((s^2 + \lambda^2)^{-m/2} [\lambda + \lambda(s + \lambda)]) \\ = O\left(\frac{\lambda}{(s^2 + \lambda^2)^{m/2}} + \frac{1}{s^{m-2}}\right).$$

Consideriamo ora la derivata $\partial^2 V(M, Q)/\partial \nu \partial \lambda$. Siccome le coordinate del punto Q sono $\eta_i = -\xi_i = -\lambda a_{im}(\Omega)$, risulta

$$\frac{\partial^2 V(M, Q)}{\partial \nu \partial \lambda} = - \Sigma a_{im}(\Omega) \frac{\partial^2 V(M, Q)}{\partial \nu \partial \eta_i}$$

e quindi si ricava dalla (36)

$$(39) \quad \frac{\partial^2 V(M, Q)}{\partial \nu \partial \lambda} = m(m-2)Z(M, Q)^{-(m+2/2)}(x_m + \xi_m)\Sigma(x_k + \xi_k) \cos(nx_k) \\ + (m-2)Z(M, Q)^{-m/2}\{-\Sigma a_{lm}(M) \cos(nx_l) \\ + \Sigma a_{lm}(\Omega)a_{il}(M)A_{ikh}(\Omega)(x_k + \xi_k) \cos(nx_l)\} + \mu$$

dove μ soddisfa, per le (37), (38), alla limitazione

$$(40) \quad \mu = O\left(\frac{\lambda}{(s^2 + \lambda^2)^{m/2}} + \frac{1}{s^{m-2}}\right).$$

Si ha poi, per le (36), (39),

$$(41) \quad \frac{\partial^2 V(M, P)}{\partial \nu \partial \lambda} + \frac{\partial^2 V(M, Q)}{\partial \nu \partial \lambda} \\ = -m(m-2)\{Z(M, P)^{-(m+2/2)} - Z(M, Q)^{-(m+2/2)}\}\xi_m \Sigma \xi_k \cos(nx_k) \\ + (m-2)\{Z(M, P)^{-m/2} - Z(M, Q)^{-m/2}\}\{\Sigma a_{lm}(M) \cos(nx_l) \\ - \Sigma a_{lm}(\Omega)a_{il}(M)A_{ikh}(\Omega)x_k \cos(nx_l)\} + \psi$$

e risulta, per le (29), (33), (36), (37), (38), (39), (40),

$$(42) \quad \psi = O\left\{\frac{\lambda}{(s^2 + \lambda^2)^{m/2}} + \frac{1}{s^{m-2}}\right\}.$$

Si ha inoltre

$$Z(M, P) - Z(M, Q) = \Sigma a_{ik}(P)(x_i - \xi_i)(x_k - \xi_k) - \Sigma A_{ik}(Q)(x_i + \xi_i)(x_k + \xi_k) \\ = \Sigma(A_{ik}(P) - A_{ik}(Q))(x_i x_k + \xi_i \xi_k) - 2\Sigma(A_{ik}(P) + A_{ik}(Q))x_i \xi_k$$

e quindi, osservando che risulta, per la (26),

$$\Sigma(A_{ik}(P) + A_{ik}(Q))x_i \xi_k = 2\lambda x_m + \lambda \Sigma(A_{ik}(P) + A_{ik}(Q) - 2A_{ik}(\Omega))x_i a_{km}(\Omega),$$

si ricava

$$Z(M, P) - Z(M, Q) = O(\lambda(s^2 + \lambda^2)).$$

Ne segue, per h intero positivo e per la (32),

$$Z(M, P)^h - Z(M, Q)^h = (Z(M, P) - Z(M, Q)) \sum_{i=0}^{h-1} Z(M, P)^i Z(M, Q)^{h-1-i} \\ = O(\lambda(s^2 + \lambda^2)^h).$$

Si ha perciò

$$(43) \quad Z(M, P)^{-h/2} - Z(M, Q)^{-h/2} = \frac{Z(M, Q)^h - Z(M, P)^h}{(Z(M, P)^{h/2} + Z(M, Q)^{h/2}) Z(M, P)^{h/2} Z(M, Q)^{h/2}} \\ = O\left(\frac{\lambda}{(s^2 + \lambda^2)^{h/2}}\right).$$

e quindi dalle (41), (42), (43) segue

$$\frac{\partial^2 V(M, P)}{\partial v \partial \lambda} + \frac{\partial^2 V(M, Q)}{\partial v \partial \lambda} = O\left(\frac{\lambda^3}{(s^2 + \lambda^2)^{m+2/2}} + \frac{\lambda}{(s^2 + \lambda^2)^{m/2}} + \frac{1}{s^{m-2}}\right) \\ = O\left(\frac{\lambda}{(s^2 + \lambda^2)^{m/2}} + \frac{1}{s^{m-2}}\right).$$

Perciò anche la terza della (22) è dimostrata.

b) Ricordiamo che, per ipotesi, i coefficienti a_{ik} , b_i , c della (1) si possono prolungare in un dominio τ' contenente τ al suo interno (in modo da risultare in τ' di classi $2H$, $1H$, H rispettivamente) e consideriamo la funzione

$$(44) \quad W(S, R) = \int_{\tau'} V(S, N) E^*_{\mathcal{N}}(V(N, R)) d\tau'$$

dove S, R sono due punti qualsiasi di τ' e si è indicato, per maggior chiarezza, con $E^*_{\mathcal{N}}$ l'operatore differenziale

$$\Sigma a_{ik}(N) (\partial^2 / \partial y_i \partial y_k) + \Sigma b^*_i(N) (\partial / \partial y_i) + c^*(N),$$

(y_1, \dots, y_m) essendo le coordinate del punto N .

Se M, P, Q indicano ancora gli stessi punti considerati in a), dimostriamo che valgono le limitazioni

$$(45) \quad \frac{\partial W(M, P)}{\partial v} = O\left(\frac{1}{s^{m-2}}\right), \\ \frac{\partial W(M, P)}{\partial \lambda} = O\left(\frac{1}{s^{m-2}}\right), \\ \frac{\partial^2 W(M, P)}{\partial v \partial \lambda} = O\left(\frac{\lambda}{(s^2 + \lambda^2)^{m/2}} + \frac{1}{s^{m-2}}\right)$$

e le analoghe ottenute scambiando P con Q .

Cominciamo col ricordare una nota proprietà della funzione $V(N, R)$. Si ha $V(N, R) = Z(N, R)^{-(m-2/2)}$ e quindi, indicate con (ξ_1, \dots, ξ_m) le coordinate di R ,

$$(46) \quad \frac{\partial V(N, R)}{\partial y_i} = -(m-2) Z(N, R)^{-m/2} \Sigma A_{ir}(R) (y_r - \xi_r),$$

$$(47) \quad \frac{\partial^2 V(N, R)}{\partial y_i \partial y_k} = m(m-2) Z(N, R)^{-(m+2/2)} \Sigma A_{ks}(R) A_{ir}(R) (y_s - \xi_s) (y_r - \xi_r) \\ - (m-2) Z(N, R)^{-m/2} A_{ik}(R),$$

da cui

$$\begin{aligned}
 (48) \quad \Sigma a_{ik}(N) \frac{\partial^2 V(N, R)}{\partial y_i \partial y_k} &= m(m-2) Z(N, R)^{-(m+2/2)} \Sigma a_{ik}(N) A_{ks}(R) A_{ir}(R) \\
 &\quad \times (y_s - \xi_s)(y_r - \xi_r) \\
 &\quad - (m-2) Z(N, R)^{-m/2} \Sigma a_{ik}(N) A_{ik}(R) \\
 &= m(m-2) Z(N, R)^{-(m+2/2)} \Sigma (a_{ik}(N) - a_{ik}(R)) \\
 &\quad \times A_{ks}(R) A_{ir}(R) (y_s - \xi_s)(y_r - \xi_r) \\
 &\quad - (m-2) Z(N, R)^{-m/2} \Sigma (a_{ik}(N) - a_{ik}(R)) A_{ik}(R).
 \end{aligned}$$

Ne seguono le note limitazioni

$$(49) \quad \Sigma a_{ik}(N) \frac{\partial^2 V(N, R)}{\partial y_i \partial y_k} = O(NR^{-(m-1)}),$$

$$(50) \quad E_N^*(V(N, R)) = O(NR^{-(m-1)}).$$

Siccome è anche

$$(51) \quad \frac{\partial V(M, N)}{\partial \nu} = O(MN^{-(m-1)}),$$

si deduce, per la (32),

$$\begin{aligned}
 \frac{\partial W(M, P)}{\partial \nu} &= \frac{\partial V(M, N)}{\partial \nu} E_N^*(V(N, P)) d\tau' \\
 &= O(MP^{-(m-2)}) = O(Z(M, P)^{-(m-2/2)}) \\
 &= O(1/s^{m-2})
 \end{aligned}$$

cioè la prima delle (45) è dimostrata.

Per studiare la derivata $\partial W(M, P)/\partial \lambda$ trasformiamo la (44) con la formula di Green. Per questo, si circondino i punti M e P con due ipersfere σ'_M, σ'_P aventi i centri rispettivamente in M, P e raggio ϵ ; siano σ' il contorno di τ', τ'_M e τ'_P i domini ipersferici limitati da σ'_M, σ'_P .

Risulta, per la formula di Green, detta ν' la conormale in N , essendo E_N l'operatore differenziale aggiunto di E_N^* ,

$$\begin{aligned}
 (52) \quad &\int_{\tau'-\tau'_M-\tau'_P} V(M, N) E_N^*(V(N, P)) d\tau' = \int_{\tau'-\tau'_M-\tau'_P} V(N, P) E_N(V(M, N)) d\tau' \\
 &- \int_{\sigma'} \{ V(M, N) \frac{\partial V(N, P)}{\partial \nu'} - V(N, P) \frac{\partial V(M, N)}{\partial \nu'} - L(N) V(M, N) V(N, P) \} d\sigma' \\
 &- \int_{\sigma'_M} \{ V(M, N) \frac{\partial V(N, P)}{\partial \nu'} - V(N, P) \frac{\partial V(M, N)}{\partial \nu'} - L(N) V(M, N) V(N, P) \} d\sigma'_M \\
 &- \int_{\sigma'_P} \{ V(M, N) \frac{\partial V(N, P)}{\partial \nu'} - V(N, P) \frac{\partial V(M, N)}{\partial \nu'} - L(N) V(M, N) V(N, P) \} d\sigma'_P.
 \end{aligned}$$

Siccome, per $\epsilon \rightarrow 0$, i due ultimi integrali tendono rispettivamente a $\mp k_m V(M, P)$, segue dalla (52)

$$(53) \quad W(M, P) = \int_{\tau'} V(N, P) E_N(V(M, N)) d\tau' \\ - \int_{\sigma'} \{V(M, N) \frac{\partial V(N, P)}{\partial \nu'} - V(N, P) \frac{\partial V(M, N)}{\partial \nu'} - L(N) V(M, N) V(N, P)\} d\sigma'$$

da cui si deduca la seconda delle (45).

Consideriamo infine la derivata $\partial^2 W(M, P)/\partial \nu \partial \lambda$. Per la (2), posto

$$(54) \quad \bar{E}^*_N = E^*_N - c^*,$$

si ricava

$$(55) \quad \frac{\partial^2 W(M, P)}{\partial \nu \partial \lambda} = \frac{\partial}{\partial \lambda} \int_{\tau'} \frac{\partial V(M, N)}{\partial \nu} \bar{E}^*_N(V(N, P)) d\tau' \\ + \int_{\tau'} \frac{\partial V(M, N)}{\partial \nu} c^*(N) \frac{\partial V(N, P)}{\partial \lambda} d\tau'$$

e risulta

$$(56) \quad \int_{\tau'} \frac{\partial V(M, N)}{\partial \nu} c^*(N) \frac{\partial V(N, P)}{\partial \lambda} d\tau' = O(MP^{-(m-2)}) = O(1/s^{m-2}).$$

Osserviamo ora che i punti M, P, Ω sono interni a τ' ; possiamo perciò considerarli interni a un dominio ipersferico τ_0 interno a τ' , con centro in Ω e raggio r , e siccome la funzione

$$(57) \quad U(M, P) = \int_{\tau' - \tau_0} V(M, N) \bar{E}^*_N(V(N, P)) d\tau'$$

è continua in Ω insieme alle sue derivate $\partial u/\partial \nu$, $\partial u/\partial \lambda$, $\partial^2 U/\partial \nu \partial \lambda$, basterà dimostrare che risulta

$$(58) \quad \frac{\partial}{\partial \lambda} \int_{\tau_0} \frac{\partial V(M, N)}{\partial \nu} \bar{E}^*_N(V(N, P)) d\tau_0 = O\left(\frac{\lambda}{(s^2 + \lambda^2)^{m/2}} + \frac{1}{s^{m-2}}\right).$$

Poniamo ora $\delta = Z(M, P)^{\frac{1}{2}}$ e facciamo il cambiamento di variabili $x_i = \delta \bar{x}_i$, $y_i = \delta \bar{y}_i$, $\xi_i = \delta \bar{\xi}_i$. Indicato con $\bar{R} = R/\delta$ il punto che corrisponde, per la trasformazione effettuata, al punto R , si ha

$$(59) \quad V(M, N) = (1/\delta^{m-2}) \{\Sigma A_{ik}(\delta \bar{N}) (\bar{x}_i - \bar{y}_i) (\bar{x}_k - \bar{y}_k)\}^{-(m-2/2)} \\ = \frac{\bar{V}(\bar{M}, \bar{N}, \delta)}{\delta^{m-2}} = \frac{\bar{Z}(\bar{M}, \bar{N}, \delta)^{-(m-2/2)}}{\delta^{m-2}}.$$

Ne segue, osservando che la normale \bar{n} e la conormale $\bar{\nu}$ in \bar{M} hanno gli stessi coseni direttori della normale n e della conormale ν in M ,

$$(60) \quad \frac{\partial V(M, N)}{\delta^{m-2}} = \frac{1}{\delta^{m-1}} \frac{\partial \bar{V}(\bar{M}, \bar{N}, \delta)}{\partial \bar{V}}$$

e inoltre, per le (46), (48) e (54),

$$(61) \quad \begin{aligned} \bar{E}^*_N(V(N, P)) &= \frac{1}{\delta^{m-1}} \{m(m-2)\bar{Z}(\bar{N}, \bar{P}, \delta)^{-(m+2/2)} \Sigma \frac{a_{ik}(\delta\bar{N}) - a_{ik}(\delta\bar{P})}{\delta} \\ &\quad \times A_{ir}(\delta\bar{P})(A_{ks}(\delta\bar{P})(\bar{y}_r - \bar{\xi}_r)(\bar{y}_s - \bar{\xi}_s) \\ &\quad - (m-2)Z(\bar{N}, \bar{P}, \delta)^{-m/2} [\Sigma \frac{a_{ik}(\delta\bar{N}) - a_{ik}(\delta\bar{P})}{\delta} \\ &\quad \times A_{ik}(\delta P) + \Sigma b^*_{ik}(\delta\bar{N})A_{ik}(\delta P)(\bar{y}_k - \bar{\xi}_k)]\} \\ &= \frac{\psi(\bar{N}, \bar{P}, \delta)}{\delta^{m-1}}. \end{aligned}$$

Ora, se $g(N)$ è una funzione di classe 2 per $\Omega N \leq r$, si ha

$$(62) \quad \frac{\partial g(\delta\bar{N})}{\partial \delta} = \Sigma \bar{y}_i \frac{\partial g(N)}{\partial y_i} = O(\Omega\bar{N}),$$

$$(63) \quad \begin{aligned} &\frac{\partial}{\partial \delta} \frac{g(\delta\bar{N}) - g(\delta\bar{P})}{\partial \delta} \\ &= \frac{(\Sigma \delta \bar{y}_i (\partial g(N)/\partial y_i) - g(N)) - (\Sigma \delta \bar{\xi}_i (\partial g(P)/\partial \xi_i) - g(P))}{\delta^2} \\ &= \frac{(\Sigma y_i (\partial g(N)/\partial y_i) - g(N)) - (\Sigma \xi_i (\partial g(P)/\partial \xi_i) - g(P))}{\delta^2} \\ &= \Sigma (\bar{y}_k - \bar{\xi}_k) \bar{I}_i (\partial^2 g(T)/\partial t_i \partial t_k) = O((\bar{\Omega}\bar{N} + \bar{\Omega}\bar{P})\bar{N}\bar{P}). \end{aligned}$$

essendo $T(t_1, \dots, t_m)$ un punto del segmento NP .

Si ha poi, per le (59), (60), (61),

$$(64) \quad \begin{aligned} \frac{\partial \bar{V}(\bar{M}, \bar{N}, \delta)}{\partial \bar{V}} &= - (m-2)\bar{Z}(\bar{M}, \bar{N}, \delta)^{-m/2} \Sigma A_{ik}(\delta\bar{N})(\bar{x}_k - \bar{y}_k)a_{il} \\ &\quad \times (\delta\bar{M}) \cos(\bar{n}\bar{x}_l) \\ &= O(\bar{M}\bar{N}^{-(m-1)}), \end{aligned}$$

$$(65) \quad \psi(\bar{N}, \bar{P}, \delta) = O(\bar{N}\bar{P}^{-(m-1)}).$$

Inoltre, se h è un intero positivo, risulta, per le (62), (63), (64),

$$(66) \quad \begin{aligned} \frac{\partial \bar{Z}(\bar{N}, \bar{P}, \delta)^{-h/2}}{\partial \delta} &= - (h/2)\bar{Z}(\bar{N}, \bar{P}, \delta)^{-(h+2/2)} \Sigma (\partial A_{ik}(\delta\bar{P})/\partial \delta)(\bar{y}_i - \bar{\xi}_i)(\bar{y}_k - \\ &= O(\bar{\Omega}P/\bar{N}\bar{P}^h) = O(\bar{\Omega}\bar{P} + \bar{\Omega}\bar{N}/\bar{N}\bar{P}^h), \end{aligned}$$

$$\begin{aligned}
 (67) \quad \frac{\partial^2 \bar{V}(\bar{M}, \bar{N}, \delta)}{\partial \bar{v} \partial \delta} &= \frac{m(m-2)}{2} \bar{Z}(\bar{M}, \bar{N}, \delta)^{-(m+2/2)} \sum \frac{\partial A_{ik}(\delta \bar{N})}{\partial \delta} \\
 &\quad \times (\bar{x}_i - \bar{y}_k)(\bar{x}_k - \bar{y}_i) A_{rs}(\delta \bar{N})(\bar{x}_r - \bar{y}_s) a_{sl}(\delta \bar{M}) \cos(\bar{n} \bar{x}_l) \\
 &\quad - (m-2) \bar{Z}(\bar{M}, \bar{N}, \delta)^{-m/2} \sum (\bar{x}_k - \bar{y}_k) \cos(\bar{n} \bar{x}_l) \\
 &\quad \times \frac{\partial A_{ik}(\delta \bar{N})}{\partial \delta} a_{il}(\delta \bar{M}) + A_{ik}(\delta \bar{N}) \frac{\partial a_{il}(\delta \bar{M})}{\partial \delta} \\
 &= O \frac{\bar{\Omega} \bar{N} + \bar{\Omega} \bar{M}}{\bar{M} \bar{N}^{m-1}}
 \end{aligned}$$

$$(68) \quad \frac{\partial \psi(\bar{N}, \bar{P}, \delta)}{\partial \delta} = O \left(\frac{\bar{\Omega} \bar{N} + \bar{\Omega} \bar{P}}{\bar{N} \bar{P}^{m-1}} \right).$$

Per la trasformazione effettuata, al dominio ipersferico τ_0 , di raggio r e centro Ω , viene a corrispondere, in \bar{S}_m , un dominio ipersferico $\bar{\tau}_0$, di raggio $\bar{r} = r/\delta$ e centro $\bar{\Omega}$. Supponiamo inoltre che \bar{M} e \bar{P} abbiano da Ω distanza $\leq r/2$; perciò sarà anche $\bar{\Omega} \bar{M} \leq \bar{r}/2$, $\bar{\Omega} \bar{P} \leq \bar{r}/2$.

Si ha poi, per le (60), (61),

$$\begin{aligned}
 9) \quad \frac{\partial}{\partial \lambda} \int_{\tau_0} \frac{\partial V(M, N)}{\partial \nu} \bar{E}^*_N(V(N, P) d\tau_0 &= \frac{\partial}{\partial \lambda} \left(\frac{1}{\delta^{m-2}} \int_{\bar{\tau}_0} \frac{\partial \bar{V}(\bar{M}, \bar{N}, \delta)}{\partial \bar{v}} \psi(\bar{N}, \bar{P}, \delta) d\bar{\tau}_0 \right) \\
 &= \left\{ -\frac{m-2}{\delta^{m-1}} \int_{\bar{\tau}_0} \frac{\partial \bar{V}(\bar{M}, \bar{N}, \delta)}{\partial \bar{v}} \psi(\bar{N}, \bar{P}, \delta) d\bar{\tau}_0 \right. \\
 &\quad \left. + \frac{1}{\delta^{m-2}} \frac{\partial}{\partial \delta} \int_{\bar{\tau}_0} \frac{\partial \bar{V}(\bar{M}, \bar{N}, \delta)}{\partial \bar{v}} \psi(\bar{N}, \bar{P}, \delta) d\bar{\tau}_0 \right\} \frac{\partial \delta}{\partial \lambda}
 \end{aligned}$$

da cui, osservando che, per il cambiamento di variabili effettuato, la distanza $\bar{M} \bar{P}$ soddisfa a una limitazione del tipo

$$0 < h \leq \bar{M} \bar{P} \leq k < \infty,$$

si ricava, per le (64), (65),

$$\begin{aligned}
 (70) \quad \int_{\bar{\tau}_0} \frac{\partial \bar{V}(\bar{M}, \bar{N}, \delta)}{\partial \bar{v}} \psi(\bar{N}, \bar{P}, \delta) d\bar{\tau}_0 \\
 = O \left(\int_{-\infty}^{\infty} d\bar{y}_1 \cdots \int_{-\infty}^{\infty} \bar{M} \bar{N}^{-(m-1)} \bar{N} \bar{P}^{-(m-1)} d\bar{y}_m \right) = O(1).
 \end{aligned}$$

Detta poi $\bar{\sigma}_0$ l'ipersfera che delimita $\bar{\tau}_0$, si ha

$$\begin{aligned}
 (71) \quad \frac{\partial}{\partial \delta} \int_{\bar{\tau}_0} \frac{\partial \bar{V}(\bar{M}, \bar{N}, \delta)}{\partial \bar{v}} \psi(\bar{N}, \bar{P}, \delta) d\bar{\tau}_0 &= \int_{\bar{\tau}_0} \frac{\partial}{\partial \delta} \left(\frac{\partial \bar{V}(\bar{M}, \bar{N}, \delta)}{\partial \bar{v}} \right) \psi(\bar{N}, \bar{P}, \delta) d\bar{\tau}_0 \\
 &\quad - \frac{r}{\delta^2} \int_{\bar{\sigma}_0} \frac{\partial \bar{V}(\bar{M}, \bar{N}, \delta)}{\partial \bar{v}} \psi(\bar{N}, \bar{P}, \delta) d\bar{\sigma}_0.
 \end{aligned}$$

e, per le (64), (65), (67), (68), risulta, ricordando che su $\bar{\sigma}_0$ è $\bar{M}\bar{N} \geq \bar{r}/2$, $\bar{P}\bar{N} \geq \bar{r}/2$,

$$(72) \quad \int_{\bar{\tau}_0} \frac{\partial}{\partial \delta} \left(\frac{\partial \bar{V}(\bar{M}, \bar{N}, \delta)}{\partial \bar{v}} \psi(\bar{N}, \bar{P}, \delta) \right) d\bar{\tau}_0 \\ = O(1/\delta \int_{-\infty}^{\infty} d\bar{y}_1 \cdots \int_{-\infty}^{\sigma_0} \bar{M} \bar{N}^{-(m-1)} \bar{N} \bar{P}^{-(m-1)} d\bar{y}_m) = O(1/\delta),$$

$$(73) \quad \frac{r}{\partial^2} \int_{\sigma_0} \frac{\partial \bar{V}(\bar{M}, \bar{N}, \delta)}{\partial \bar{v}} \psi(\bar{N}, \bar{P}, \delta) d\bar{\sigma}_0 = O(1/\delta^2 \int_{\bar{\sigma}_0} \delta^{2(m-1)} d\bar{\sigma}_0) = O(\delta^{m-3}).$$

Dalle (69), (70), (71), (72), (73) si ricava perciò

$$(74) \quad \frac{\partial}{\partial \lambda} \int_{\bar{\tau}_0} \frac{\partial V(M, N)}{\partial v} \bar{E}^*_N(V(N, P)) d\bar{\tau}_0 = O\left(\frac{1}{\partial^{m-1}} \frac{\partial \delta}{\partial \lambda}\right).$$

Ora si ha

$$\delta^2 = \Sigma A_{ik}(P) (x_i - \xi_i) (x_k - \xi_k)$$

e quindi, per la (26),

$$(75) \quad 2\delta \frac{\partial \delta}{\partial \lambda} = \Sigma \frac{\partial A_{ik}(P)}{\partial \lambda} (x_i - \xi_i) (x_k - \xi_k) + 2\lambda \Sigma A_{ik}(P) a_{im}(\Omega) a_{km}(\Omega) \\ - 2\Sigma A_{ik}(P) x_i a_{km}(\Omega) = O(s^2 + \lambda)$$

come segue dalle (27), (28), (29), (31).

Dalle (55), (56), (57), (74), (75), segue allora la terza delle (45).

c) Dimostriamo ora, valendoci delle (22) e (45), le analoghe limitazioni per la soluzione fondamentale della (4).

Come è noto, E. E. Levi¹⁰ ha posto tale funzione nella forma

$$(76) \quad F(N, P) = V(N, P) + \int_{\tau'} V(N, T) \phi(T, P) d\tau' + \psi(N, P)$$

con ϕ, ψ funzioni opportune. Per determinarle si impone alla $F(N, P)$ di soddisfare, come funzione di N , all'equazione (4).

Detta σ_m la misura ipersuperficiale dell'ipersfera di raggio 1 in S_m , si ricava l'equazione integrale di Fredholm

$$(77) \quad \sigma_m \phi(N, P) = \int_{\tau'} E^*_N(V(N, T)) \phi(T, P) d\tau' \\ + E^*_N(V(N, P)) + E^*_N(\psi(N, P)).$$

¹⁰ G. Ascoli, *loc. cit.*², p. 72.

Se il determinante della (77) è $\neq 0$, si prende $\psi(N, P) \equiv 0$; in caso contrario, per fare in modo che la (77) abbia soluzione, il Levi pone

$$\psi(N, P) = \sum u_i(N) \eta_i(P)$$

dove le $u_i(N)$ sono opportune funzioni di classe 2 in τ' e le $\eta_i(P)$ risultano combinazioni lineari delle autofunzioni $\psi_1(P), \dots, \psi_k(P)$ del nucleo associato $(1/\sigma_m)E^*_T(V(T, N))$:

$$(78) \quad \psi_i(P) = 1/\sigma_m \int_{\tau'} E^*_T(V(T, P)) \psi_i(T) d\tau'.$$

Ciò premesso, osserviamo che dalla (77), posto

$$K(N, T) = (1/\sigma_m)E^*_N(V(N, T)), \quad \chi(N, P) = (1/\sigma_m)E^*_N(\psi(N, P)),$$

segue

$$(79) \quad \phi(N, P) - K(N, P) = \int_{\tau'} K(N, T) (\phi(T, P) - K(T, P)) d\tau' \\ + K_2(N, P) + \chi(N, P)$$

essendo $K_2(N, P)$ il primo nucleo iterato di $K(N, P)$.

Detto $H(N, T)$ il nucleo risolvante della (79), risulta allora

$$(80) \quad \phi(N, P) = K(N, P) + K_2(N, P) + \chi(N, P) \\ + \int_{\tau'} H(N, T) (K_2(T, P) + \chi(T, P)) d\tau'.$$

Ora, ragionando come in b), si ricava

$$(81) \quad K_2(T, P) = O(TP^{-(m-2)}), \\ \frac{\partial K_2(T, P)}{\partial \lambda} = O(TP^{-(m-1)})$$

ed è inoltre

$$(82) \quad H(N, T) = O(NT^{-(m-1)}).$$

Siccome poi, per la (78), le funzioni $\psi_i(P)$ soddisfano anche all'equazione

$$\psi_i(P) = \int_{\tau'} K_{m+1}(T, P) \psi_i(T) d\tau'$$

e risultano perciò di classe 1 (essendo il nucleo iterato $K_{m+1}(T, P)$ dotato di derivate parziali rispetto alle coordinate di P continue per $P \neq T$, limitate in τ') anche la funzione $\chi(N, P)$ è di classe 1, come funzione di P .

La tesi ¹¹ segue allora dalle (22), (45) e dalle (79), (80), (81), (82).

2. In quel che segue faremo le seguenti ipotesi, largamente verificate nei casi pratici, sulla natura del dominio limitato τ e delle funzioni A , B che compaiono nelle (15), (16).

Supporremo che il contorno σ abbia misura ipersuperficiale finita e risulta di classe 2 in ogni suo punto, esclusi al più i punti di un insieme chiuso χ di misura ipersuperficiale nulla. Ammetteremo inoltre che valga in τ la formula di Green per ogni coppia u , w di funzioni di classe 2. Si noti che, nelle ipotesi poste, la connessione ipersuperficiale di τ può anche essere infinita.

Ciò premesso, diremo che σ , A , B appartengono all'insieme (α) se σ soddisfa alle condizioni ora indicate e se A , B sono sommabili su σ .

Converrà inoltre studiare un altro caso, più comune nelle applicazioni, considerando un secondo insieme, (β) , contenuto in (α) .

Anzitutto, supporremo $A(M)$ continua in tutti i punti di σ , $B(M)$ continua in tutti i punti di $\sigma - \chi$ e tale che l'integrale

$$(83) \quad J_1(N) = \int_{\sigma} |B(M)| MN^{-(m-2)} d\sigma, \quad \text{per } m > 2,$$

$$J_1(N) = \int_{\sigma} |B(M) \log MN| d\sigma, \quad \text{per } m = 2,$$

considerato per N variabile su tutto σ , sia ivi funzione continua.

Supporremo inoltre che gli integrali

$$(84) \quad \begin{aligned} J_2(R) &= \int_{\sigma} MR^{-(m-2)} d\sigma, & J_3(R) &= \int_{\sigma} \left| \frac{\partial}{\partial n} MR^{-(m-2)} \right| d\sigma, \\ & & & \text{per } m > 2, \\ J_2(R) &= \int_{\sigma} |\log MR| d\sigma, & J_3(R) &= \int_{\sigma} \left| \frac{\partial}{\partial n} \log MR \right| d\sigma, \\ & & & \text{per } m = 2, \end{aligned}$$

si mantengano limitati al variare del punto R in tutto lo spazio S_m .

Il significato geometrico della limitatezza degli integrali $J_3(R)$ è evi-

¹¹ E. E. Levi, "Sulle equazioni lineari totalmente ellittiche alle derivate parziali," *Rendiconti del circolo Matematico di Palermo*, vol. 24 (1907), pp. 275-317. Vedasi anche, dello stesso Autore, "I problemi dei valori al contorno per le equazioni lineari totalmente ellittiche alle derivate parziali," *Memorie Società Italiana dei XL*, vol. 16 (1910), pp. 3-111. La funzione $V(M, P)$ data dalle (20), (21) è quella ora generalmente considerata (G. Ascoli, *loc. cit.*², p. 72) e risulta leggermente diversa da quella introdotta dal Levi.

dente quando si ricordi che, detto $d\omega_{M,R}$ l'angolo secondo il quale è visto dal punto R l'elemento $d\sigma$ relativo al punto M , risulta, come è noto,

$$J_3(R) = \int_{\sigma} |d\omega_{M,R}|.$$

Ciò premesso, indicheremo con Γ_a la totalità degli integrali della (1) soddisfacenti alle seguenti condizioni:

a) $u(P)$ è dato nell'interno di τ dalla formula (16) dove σ , A , B appartengono all'insieme α e le funzioni A , B soddisfano all'equazione (15);

b) risulta, in quasi tutti i punti M di σ ,

$$(85) \quad \lim_{P \rightarrow M} u(P) = A(M),$$

$$\lim_{P \rightarrow M} \frac{\partial u(P)}{\partial \nu} = B(M),$$

dove $P \rightarrow M$ lungo la conormale ν a σ , condotta dal punto M stesso.

Indicheremo poi con Γ_β la totalità degli integrali della (1) che si ottengono da Γ_a imponendo, come ulteriori condizioni, che σ , A , B , appartengano all'insieme (β) , che la prima delle (85) valga in tutti i punti di σ , comunque $P \rightarrow M$, e, infine, che la seconda delle (85) valga in tutti i punti di $\sigma - \chi$ per $P \rightarrow M$ lungo la conormale ν .

3. Dimostriamo ora il seguente teorema fondamentale.

I. Se le funzioni $A(M)$, $B(M)$ soddisfano all'equazione (15) in tutti i punti Q del dominio τ' esterni a τ , la funzione $u(P)$ definita entro τ dalla (16) appartiene all'insieme Γ_a o Γ_β a seconda che σ , A , B appartengano all'insieme (α) o (β) .

Dimostrazione. Supporremo $m > 2$, la dimostrazione per $m = 2$ essendo del tutto analoga.

Siccome la funzione f è di classe H in τ , la $u(P)$ definita dalla (16) soddisfa all'equazione (1), qualunque siano le funzioni $A(M)$, $B(M)$ sommabili su σ .

a) Ammettiamo dapprima che σ , A , B appartengano all'insieme (β) . Sia M_0 un punto qualsiasi di σ . Siccome risulta, per le (13), (14),

$$k_m \equiv \int_{\sigma} \left\{ \frac{\partial F(M, P)}{\partial \nu} - L(M)F(M, P) \right\} d\sigma - \int_{\tau} c(M)F(M, P) d\tau,$$

$$0 \equiv \int_{\sigma} \left\{ \frac{\partial F(M, Q)}{\partial \nu} - L(M)F(M, Q) \right\} d\sigma - \int_{\tau} c(M)F(M, Q) d\tau,$$

si ricava dalle (15), (16)

$$(86) \quad k_m(u(P) - A(M_0)) = \int_{\sigma} \{ (A(M) - A(M_0)) \left[\frac{\partial F(M, P)}{\partial \nu} - \frac{\partial F(M, Q)}{\partial \nu} \right. \right. \\ \left. \left. - L(M) (F(M, P) - F(M, Q)) \right] - B(M) (F(M, P) - F(M, Q)) \} d\sigma \\ - \int_{\tau} (f(M) - A(M_0) c(M)) (F(M, P) - F(M, Q)) d\tau$$

e, per essere $F(M, N) = O(MN^{-(m-2)})$, risulta

$$(87) \quad \lim_{(P, Q) \rightarrow M_0} \int_{\tau} (f(M) - A(M_0) c(M)) (F(M, P) - F(M, Q)) d\tau = 0.$$

Dimostriamo ora che si ha

$$(88) \quad \lim_{(P, Q) \rightarrow M_0} \int_{\sigma} B(M) (F(M, P) - F(M, Q)) d\sigma = 0.$$

Sia K una costante positiva tale che risulti

$$(89) \quad |F(M, R)| \leq KMR^{-(m-2)}$$

qualunque siani i punti M, R di τ' .

Prefissato $\epsilon > 0$ si determini poi, in corrispondenza, $\delta > 0$ in modo che, detta σ_{δ} la parte di σ contenuta nell'ipersfera (M_0, δ) , di centro M_0 e raggio δ , risulti

$$(90) \quad \int_{\sigma_{\delta}} |B(M)| MM_0^{-(m-2)} d\sigma_{\delta} \leq \epsilon/K.$$

Si ha poi, se M_1 è un punto di σ_{δ} , per la continuità dell'integrale $J_1(N)$ definito dalla (83),

$$\lim_{M_1 \rightarrow M_0} \int_{\sigma_{\delta}} |B(M)| MM_1^{-(m-2)} d\sigma_{\delta} = \int_{\sigma_{\delta}} |B(M)| MM_0^{-(m-2)} d\sigma_{\delta}$$

e quindi si può determinare $\eta \leq \delta$ in modo che, per $M_0 M_1 \leq \eta$, risulti, per le (89), (90),

$$\int_{\sigma_{\delta}} |B(M)| MM_1^{-(m-2)} d\sigma_{\delta} < 2\epsilon/\bar{K}$$

e, a maggior ragione,

$$(91) \quad \int_{\sigma_{\eta}} |B(M)| MM_1^{-(m-2)} d\sigma_{\eta} < 2\epsilon/K.$$

Sia ora R un punto qualsiasi di τ' , M_R un punto di σ_{η} situato alla minima distanza da R . Siccome è

$$MM_R \leq MR + RM_R \leq 2MR$$

risulta, per le (89), (91),

$$(92) \quad \int_{\sigma\eta} |B(M)F(M, R)| d\sigma\eta \leq K \int_{\sigma\eta} |B(M)| MR^{-(m-2)} d\sigma\eta \\ \leq 2^{m-2}K \int_{\sigma\eta} |B(M)| MM_R^{-(m-2)} d\sigma\eta < 2^{m-1}\epsilon.$$

Si ha poi

$$\int_{\sigma} B(M) (F(M, P) - F(M, Q)) d\sigma \leq \int_{\sigma\eta} |B(M)| (|F(M, P)| \\ + |F(M, Q)|) d\sigma\eta \\ + \left| \int_{\sigma-\sigma\eta} B(M) (F(M, P) - F(M, Q)) d\sigma \right|$$

e il terzo integrale a secondo membro è infinitesimo per $(P, Q) \rightarrow M_0$. Dalla (92) segue allora la (88).

Proviamo ora che risulta

$$(93) \quad \lim_{(P, Q) \rightarrow M_0} \int_{\sigma} (A(M) - A(M_0)) \left[\frac{\partial F(M, Q)}{\partial \nu} - \frac{\partial F(M, Q)}{\partial \nu} \right. \\ \left. - L(M) (F(M, P) - F(M, Q)) \right] d\sigma = 0.$$

Posto $\rho = MR$ e indicate con (ξ_1, \dots, ξ_m) le coordinate del punto R , si ha, per la (24),

$$(94) \quad \frac{\partial F(M, R)}{\partial \nu} = -(m-2)Z(M, R)^{-m/2} \Sigma(x_k - \xi_k) \cos(nx_k) + \omega(M, R)$$

con

$$(95) \quad \omega(M, R) = O(MR^{-(m-2)}).$$

Siccome è

$$\Sigma(x_k - \xi_k) \cos(nx_k) = \rho(\partial \rho / \partial n),$$

si ricava perciò dalle (92), (95),

$$(96) \quad \frac{\partial F(M, R)}{\partial \nu} = O\left(\left| \frac{\partial \rho^{-(m-2)}}{\partial n} \right| + \rho^{-(m-2)}\right)$$

e quindi, per la limitatezza degli integrali $J_2(R)$, $J_3(R)$ definiti dalla (84) e per le (89), (96),

$$(97) \quad \int_{\sigma} \left(\left| \frac{\partial F(M, R)}{\partial \nu} \right| + |L(M)F(M, R)| \right) d\sigma \leq H$$

con H costante positiva.

Prefissato $\epsilon > 0$ si determini ora, in corrispondenza $\gamma > 0$ in modo che, per $M_0 M \leq \gamma$, risulti $|A(M) - A(M_0)| < \epsilon/H$.

Ne segue, per la (97),

$$(98) \quad \left| \int_{\sigma} (A(M) - A(M_0)) \left[\frac{\partial F(M, P)}{\partial \nu} - \frac{\partial F(M, Q)}{\partial \nu} - L(M) (F(M, P) - F(M, Q)) \right] d\sigma \right| \\ < \left| \int_{\sigma - \sigma\gamma} (A(M) - A(M_0)) \left[\frac{\partial F(M, P)}{\partial \nu} - \frac{\partial F(M, Q)}{\partial \nu} - L(M) (F(M, P) - F(M, Q)) \right] d\sigma \right| + 2\epsilon$$

e quindi, essendo l'integrale a secondo membro infinitesimo per $(P, Q) \rightarrow M_0$, la (93) è dimostrata.

Dalle (87), (88), (93) risulta poi

$$(99) \quad \lim_{P \rightarrow M_0} u(P) = A(M_0).$$

Proviamo ora che, se M_0 è un punto di $\sigma - \chi$, detta λ la conormale in M_0 a σ , si ha

$$(100) \quad \lim_{P \rightarrow M_0} \frac{\partial u(P)}{\partial \lambda} = B(M_0),$$

dove $P \rightarrow M_0$ lungo la conormale in M_0 .

Portiamo, con un cambiamento di coordinate, in M_0 l'origine Ω delle coordinate e assumiamo come asse x_m la normale a σ in M_0 orientata verso l'interno di τ .

Si ha allora, per la (26), applicando le formule (13), (14) alla funzione $x_m/a_{mm}(M_0)$ e supposto P sulla conormale in M_0 ,

$$k_m \lambda = \int_{\sigma} \left\{ \frac{x_m}{a_{mm}(M_0)} \left(\frac{\partial F(M, P)}{\partial \nu} - L(M) F(M, P) \right) - \frac{a_{mm}(M)}{a_{mm}(M_0)} F(M, P) \right\} d\sigma \\ - \int_{\tau} \frac{b_m(M) + x_m c(M)}{a_{mm}(M_0)} F(M, P) d\tau \\ 0 \equiv \int_{\sigma} \left\{ \frac{x_m}{a_{mm}(M_0)} \left(\frac{\partial F(M, Q)}{\partial \nu} - L(M) F(M, Q) \right) - \frac{a_{mm}(M)}{a_{mm}(M_0)} F(M, Q) \right\} d\sigma \\ - \int_{\tau} \frac{b_m(M) + x_m c(M)}{a_{mm}(M_0)} F(M, Q) d\tau.$$

Ne segue

$$\begin{aligned}
 k_m(u(P) - A(M_0) - \lambda B(M_0)) = & \int_{\sigma} \left\{ \left(A(M) - A(M_0) \right. \right. \\
 & \left. \left. - \frac{x_m B(M_0)}{a_{mm}(M_0)} \right) \left[\frac{\partial F(M, P)}{\partial \nu} + \frac{\partial F(M, Q)}{\partial \nu} \right. \right. \\
 & \left. \left. - L(M)(F(M, P) + F(M, Q)) \right] \right. \\
 & \left. - \left(B(M) - \frac{a_{mm}(M) B(M_0)}{a_{mm}(M_0)} \right) (F(M, P) + F(M, Q)) \right\} d\sigma \\
 & - \int_{\tau} \left(f(M) - c(M) A(M_0) - \frac{(b_m(M) + x_m c(M)) B(M_0)}{a_{mm}(M_0)} \right) \\
 & (F(M, P) + F(M, Q)) d\tau
 \end{aligned}$$

e quindi, assunto come punto Q il simmetrico di P rispetto a M_0 ,

$$\begin{aligned}
 (101) \quad k_m \left(\frac{\partial u(P)}{\partial \lambda} - B(M_0) \right) = & \int_{\sigma} \left\{ \left(A(M) - A(M_0) - \frac{x_m B(M_0)}{a_{mm}(M_0)} \right) \right. \\
 & \left[\frac{\partial^2 F(M, P)}{\partial \nu \partial \lambda} + \frac{\partial^2 F(M, Q)}{\partial \nu \partial \lambda} \right. \\
 & \left. - L(M) \left(\frac{\partial F(M, P)}{\partial \lambda} + \frac{\partial F(M, Q)}{\partial \lambda} \right) \right] - \left(B(M) - \frac{a_{mm}(M) B(M_0)}{a_{mm}(M_0)} \right) \\
 & \left(\frac{\partial F(M, P)}{\partial \lambda} + \frac{\partial F(M, Q)}{\partial \lambda} \right) \left. \right\} d\sigma \\
 & - \int_{\tau} \left(f(M) - c(M) A(M_0) - \frac{(b_m(M) + x_m c(M)) B(M_0)}{a_{mm}(M_0)} \right) \\
 & \left(\frac{\partial F(M, P)}{\partial \lambda} + \frac{\partial F(M, Q)}{\partial \lambda} \right) d\tau.
 \end{aligned}$$

Osserviamo ora che risulta, uniformemente, al variare di M in un insieme chiuso non contenente M_0 ,

$$(102) \quad \lim_{\lambda \rightarrow 0} \left(\frac{\partial F(M, P)}{\partial \lambda} + \frac{\partial F(M, Q)}{\partial \lambda} \right) = \lim_{\lambda \rightarrow 0} \left(\frac{\partial^2 F(M, P)}{\partial \nu \partial \lambda} + \frac{\partial^2 F(M, Q)}{\partial \nu \partial \lambda} \right) = 0.$$

Per il lemma I si possono determinare due numeri positivi h, k in modo che, per $M_0 M \leq h$, $0 < \lambda \leq h$, risulti

$$\begin{aligned}
 & \left| \frac{\partial^2 F(M, P)}{\partial \nu \partial \lambda} + \frac{\partial^2 F(M, Q)}{\partial \nu \partial \lambda} \right| + (1 + |L(M)|) \left| \frac{\partial F(M, P)}{\partial \lambda} + \frac{\partial F(M, Q)}{\partial \lambda} \right| \\
 & \leq k \left(\frac{\lambda}{(s^2 + \lambda^2)^{m/2}} + \frac{1}{s^{m-2}} \right)
 \end{aligned}$$

e quindi

$$\begin{aligned}
 (103) \quad & \int_{\sigma_n} \left\{ \left| \frac{\partial^2 F(M, P)}{\partial v \partial \lambda} + \frac{\partial^2 F(M, Q)}{\partial v \partial \lambda} \right| + (1 + |L(M)|) \right. \\
 & \quad \left. \left| \frac{\partial F(M, P)}{\partial \lambda} + \frac{\partial F(M, Q)}{\partial \lambda} \right| \right\} d\sigma_n \\
 & \leq k' \int_0^h s^{m-2} \left(\frac{\lambda}{(s^2 + \lambda^2)^{m/2}} + \frac{1}{s^{m-2}} \right) ds < k' \left(\int_0^\infty \frac{\lambda}{s^2 + \lambda^2} ds + h \right) \\
 & = k'(\pi/2 + h) = k''
 \end{aligned}$$

con k', k'' costanti positive.

Prefissato $\epsilon > 0$, si determini in corrispondenza δ , con $0 < \delta \leq h$, in modo che per $M_0 M \leq \delta$ risulti

$$\begin{aligned}
 (104) \quad & \left| A(M) - A(M_0) - \frac{x_m B(M_0)}{a_{mm}(M_0)} \right| < \epsilon/k'', \\
 & \left| B(M) - B(M_0) \frac{a_{mm}(M)}{a_{mm}(M_0)} \right| < \epsilon/k''.
 \end{aligned}$$

Dalle (101), (103), (104) segue allora

$$\begin{aligned}
 (105) \quad & |k_m(\partial u(P)/\partial \lambda - B(M_0))| \leq \epsilon \\
 & + \left| \int_{\sigma-\sigma\delta} \left\{ \left(A(M) - A(M_0) - \frac{x_m B(M_0)}{a_{mm}(M_0)} \right) \left[\frac{\partial^2 F(M, P)}{\partial v \partial \lambda} + \frac{\partial^2 F(M, Q)}{\partial v \partial \lambda} \right. \right. \right. \\
 & \quad \left. \left. - L(M) \left(\frac{\partial F(M, P)}{\partial \lambda} + \frac{\partial F(M, Q)}{\partial \lambda} \right) \right] - \left(B(M) - \frac{a_{mm}(M) B(M_0)}{a_{mm}(M_0)} \right) \right. \right. \\
 & \quad \left. \left. \left(\frac{\partial F(M, P)}{\partial \lambda} + \frac{\partial F(M, Q)}{\partial \lambda} \right) \right\} d\sigma \right| \\
 & + \left| \int_\tau \left(f(M) - c(M) A(M_0) - \frac{(b_m(M) + x_m c(M)) B(M_0)}{a_{mm}(M_0)} \right) \right. \\
 & \quad \left. \left(\frac{\partial F(M, P)}{\partial \lambda} + \frac{\partial F(M, Q)}{\partial \lambda} \right) d\tau \right|.
 \end{aligned}$$

Per la (102) e per le limitazioni, valide in tutti i punti M di τ ,

$$\begin{aligned}
 \left| \frac{\partial F(M, P)}{\partial \lambda} \right| & \leq KMP^{-(m-1)}, \\
 \left| \frac{\partial F(M, Q)}{\partial \lambda} \right| & \leq KMQ^{-(m-1)},
 \end{aligned}$$

(dove K è una costante positiva) gli integrali a secondo membro della (105) sono infinitesimo con λ e quindi dalla (105) segue la (100).

b) Supponiamo ora che σ , A , B appartengano all'insieme (α) . In tal caso ricordiamo che risulta quasi ovunque su $\sigma - \chi$ e quindi su σ , per la sommabilità di A , B ,

$$(106) \quad \lim_{\delta \rightarrow 0} (1/\delta^{m-1} \int_{\sigma\delta} \{|A(M) - A(M_0)| + |B(M) - B(M_0)|\} d\sigma_\delta = 0.$$

Sia M_0 un punto di $\sigma - \chi$ in cui vale la (106). Sarà perciò anche, in M_0 ,

$$(107) \quad \lim_{\delta \rightarrow 0} \int_{\sigma\delta} \left\{ \left| A(M) - A(M_0) - \frac{x_m B(M_0)}{a_{mm}(M_0)} \right| + \left| B(M) - \frac{a_{mm}(M) B(M_0)}{a_{mm}(M_0)} \right| \right\} d\sigma_\delta = 0.$$

Prefissato $\epsilon > 0$ e posto

$$\psi(t) = \int_{\sigma t} \{|A(M) - A(M_0)| + |B(M) - B(M_0)|\} d\sigma_t,$$

si determini η , con $0 < \eta \leq \epsilon$, in modo che, per $0 \leq t \leq \eta$, risulti

$$0 \leq \psi(t) \leq \epsilon t^{m-1}.$$

Si ricava allora, per il lemma I, indicata con k_1 una costante positiva indipendente da ϵ e supposti P e Q sulla conormale in M_0 (sicchè risulterà $|F(M, P)| + |F(M, Q)| = 0(t^{-(m-2)})$),

$$(108) \quad \left| \int_{\sigma\eta} \left\{ (A(M) - A(M_0)) \left[\frac{\partial F(M, P)}{\partial \nu} - \frac{\partial F(M, Q)}{\partial \nu} \right] - L(M) \Delta F(M, P) - F(M, Q) \right\} - B(M) [F(M, P) - (F(M, Q))] d\sigma_\eta \right| \\ \leq k_1 \left[\int_0^\eta \left(\frac{\lambda}{(t^2 + \lambda^2)^{m/2}} + \frac{1}{t^{m-2}} \right) \psi'(t) dt + \eta |B(M_0)| \right] \\ = k_1 \left[\psi(\eta) \left(\frac{\lambda}{(\eta^2 + \lambda^2)^{m/2}} + \frac{1}{\eta^{m-2}} \right) + \int_0^\eta \psi(t) \left(\frac{mt\lambda}{(t^2 + \lambda^2)^{m+2/2}} + \frac{m-2}{t^{m-1}} \right) dt + \eta |B(M_0)| \right] \\ < \epsilon k_1 \left[\frac{\eta^{m-1}\lambda}{(\eta^2 + \lambda^2)^{m/2}} + \eta + \int_0^\eta \frac{mt^m\lambda}{(t^2 + \lambda^2)^{m+2/2}} + m-2 dt + |B(M_0)| \right] < \epsilon k_1 (1 + \eta + m\pi/2) + (m-2)\eta + |B(M_0)|.$$

Dalla (108) segue poi, per la (86),

$$\lim_{P \rightarrow M_0} u(P) = A(M_0),$$

dove $P \rightarrow M_0$ lungo la conormale λ .

Poniamo ora

$$\omega(t) = \int_{\sigma_t} \left\{ \left| A(M) - A(M_0) - \frac{x_m B(M_0)}{a_{mm}(M_0)} \right| + \left| B(M) - \frac{a_{mm}(M) B(M_0)}{a_{mm}(M_0)} \right| \right\} d\sigma_t$$

e determiniamo $\xi > 0$ in modo che, per $0 \leq t \leq \xi$, risulti, per la (107),

$$0 \leq \omega(t) \leq \epsilon t^{m-1}.$$

Si ricava allora, per il lemma I, indicata con h_1 una costante positiva indipendente da ϵ e supposti P, Q sulla conormale in M_0 ed equidistanti da tale punto,

$$\begin{aligned} & \left| \int_{\sigma_\xi} \left\{ \left(A(M) - A(M_0) - \frac{x_m B(M_0)}{a_{mm}(M_0)} \right) \left[\frac{\partial^2 F(M, P)}{\partial \nu \partial \lambda} + \frac{\partial^2 F(M, Q)}{\partial \nu \partial \lambda} \right. \right. \right. \\ & \quad \left. \left. - L(M) \left(\frac{\partial F(M, P)}{\partial \lambda} + \frac{\partial F(M, Q)}{\partial \lambda} \right) \right] - \left(B(M) - \frac{a_{mm}(M) B(M_0)}{a_{mm}(M_0)} \right) \right. \right. \\ & \quad \left. \left. \left(\frac{\partial F(M, P)}{\partial \lambda} + \frac{\partial F(M, Q)}{\partial \lambda} \right) \right\} d\sigma_\xi \right| \\ & \leq h_1 \int_0^\xi \left\{ \frac{\lambda}{(t^2 + \lambda^2)^{m/2}} + \frac{1}{t^{m-2}} \right\} \omega'(t) dt \end{aligned}$$

e quindi, ragionando come precedentemente, si deduce

$$\lim_{P \rightarrow M_0} \frac{\partial u(P)}{\partial \lambda} = B(M_0)$$

dove $P \rightarrow M_0$ lungo la sonormale λ in tale punto.

4. Dal teorema I segue che l'equazione (15) *caratterizza* tutte le coppie A, B di funzioni coincidenti rispettivamente con i valori assunti su σ da un integrale u della (1), appartenente all'insieme Γ_α o Γ_β , e dalla sua derivata conormale $\partial u / \partial \nu$.

Essa può inoltre essere utilizzata per risolvere non solo i problemi di Dirichlet e di Neumann ma anche il problema misto (che li comprende come casi particolari) ed il problema di Cauchy (in grande).

Supponiamo infatti assegnata su σ una combinazione lineare delle funzioni u e $\partial u / \partial \nu$:

$$(109) \quad hu + k(\partial u / \partial \nu) = C,$$

con h, k, C funzioni assegnate, le prime due non contemporaneamente nulle, escluso al più un insieme γ di misura ipersuperficiale nulla. Ammettiamo inoltre che $\sigma - \gamma$ si possa dividere in due insiemi, σ_1 e σ_2 , in modo che, in σ_1 , la funzione h/k sia limitata e la C/k sommabile, in σ_2 sia limitata la funzione k/h e la C/h sommabile.

Poniamo poi nella (109) A, B in luogo di $u, \partial u / \partial \nu$ e supponiamo che la funzione $u(P)$ appartenga all'insieme Γ_α (in modo simile si può studiare il caso in cui $u(P)$ appartiene all'insieme Γ_β).

Sarà, su σ_1 ,

$$(110) \quad B = - (h/k) A + C/k$$

e, su σ_2 ,

$$(111) \quad A = - (k/h) B + C/h.$$

Introdotta la successione $\{v_r\}$ data dalla (17), risulta, per le (18), (110), (111)

$$\begin{aligned} (112) \quad & \int_{\sigma_0} A \{ (\partial v_r / \partial \nu) - v_r (L - (h/k)) \} d\sigma_1 \\ & - \int_{\sigma_2} B \left\{ \frac{k}{h} \frac{\partial v_r}{\partial \nu} + v_r (1 - L(k/h)) \right\} d\sigma_2 \\ & = \int_{\sigma_1} (C/k) v_r d\sigma_1 - \int_{\sigma_2} (C/h) \{ (\partial v_r / \partial \nu) - L v_r \} d\sigma_2 + \int_{\tau} f v_r d\tau. \end{aligned}$$

Se definiamo perciò un vettore \mathbf{G} di componenti A in σ_1 , B in σ_2 e una successione di vettori $\{\omega_r\}$, di componenti $\{(\partial v_r / \partial \nu) - v_r (L - (h/k))\}$ in σ_1 , $-\{ \frac{k}{h} \frac{\partial v_r}{\partial \nu} + v_r (1 - L(k/h)) \}$ in σ_2 , e indichiamo con c_r il secondo membro della (112), ricaviamo il sistema di Fischer-Riesz

$$(113) \quad (\mathbf{G}, \omega_r) = c_r$$

cioè risultano noti coefficienti di Fourier dell'incognito vettore \mathbf{G} rispetto alla successione $\{\omega_r\}$.

Viceversa, sia \mathbf{G} una soluzione del sistema (113). Si conoscono allora le funzioni A su σ_1 , B su σ_2 ; definite poi, su σ_1 e σ_2 rispettivamente, le B, A mediante le (110) e (111), dalla (112), eliminando C con le (110), (111), segue che la coppia A, B soddisfa al sistema (18) e quindi all'equivalente equazione (15). Per il teorema I la funzione $u(P)$ data dalla (16) è un integrale della (1) tale che risulti, su σ , $u = A, \partial u / \partial \nu = B$. Per le (110), (111) è allora verificata su σ la (109).

Inoltre, se per il problema considerato vale, nell'insieme Γ_a , il teorema di unicità, la successione $\{\omega_r\}$ risulta chiusa rispetto alla totalità dei vettori \mathbf{G} di componenti sommabili; se tale teorema non vale, le autosoluzioni sono caratterizzate dall'essere i corrispondenti vettori \mathbf{G} ortogonali su σ alla successione $\{\omega_r\}$.

Infine, se il problema non ammette soluzione, nemmeno il sistema (113) ammette soluzione.

Dedotta poi dalla successione $\{\omega_r\}$ una equivalente successione $\{\zeta_r\}$ ortogonale e normale, il sistema (113) equivale al sistema

$$(114) \quad (\mathbf{G}, \zeta_r) = d_r$$

dove le d_r sono costanti note. Perciò condizione necessaria e sufficiente affinché il problema ammetta soluzione, il vettore \mathbf{G} risultando di norma integrale finita, è che converga la serie $\sum d_r^2$.

In modo analogo si ragiona per il problema di Cauchy, nel quale supporremo assegnati su una parte, σ_1 , di σ i valori di u , e $\partial u / \partial \nu$. Posto $\sigma_2 = \sigma - \sigma_1$ e, al solito, su σ , $u = A$, $\partial u / \partial \nu = B$, si ricava dal sistema (18) il sistema

$$(115) \quad \int_{\sigma_2} \{A((\partial v_r / \partial \nu) - L v_r) - B v_r\} d\sigma_2 \\ = - \int_{\sigma_1} \{A((\partial v_r / \partial \nu) - L v_r) - B v_r\} d\sigma_1 + \int_{\tau} f v_r d\tau = c_r,$$

con c_r costanti note. Considerate poi le A, B , in σ_2 , come componenti di un vettore \mathbf{G} , le $((\partial v_r / \partial \nu) - L v_r), -v_r$ come componenti di un vettore ω_r , il sistema (115) diventa

$$\{\mathbf{G}, \omega_r\} = c_r$$

e si possono ripetere le considerazioni svolte a proposito dei sistemi (113), (114).

II. Secondo procedimento di integrazione.

1. Dimostriamo due lemmi dei quali faremo uso in seguito.

LEMMA II. Sia $U(M, R)$ una funzione definita nel dominio τ' contenente τ nel suo interno, la quale soddisfi, uniformemente al variare di M, R in τ' , alle limitazioni

$$\begin{aligned}
 U(M, R) &= O(MR^{-(m-2)}), \\
 \frac{\partial U(M, R)}{\partial x_i} &= O(MR^{-(m-1)}), \\
 \frac{\partial^2 U(M, R)}{\partial x_i \partial x_k} &= O(MR^{-m}).
 \end{aligned}$$

dove (x_1, \dots, x_m) sono le coordinate di M .

Indicato con d il diametro del dominio τ' e posto, per r intero positivo, $M, N(y_1, \dots, y_m)$ in τ' ,

$$\begin{aligned}
 \phi_r(N, M) &= 1/d^m (r/\pi)^{m/2} \{ (1 - ((x_1 - y_1)^2/d^2)) \cdots \\
 &\quad \times (1 - ((x_m - y_m)^2/d^2)) \}^2,
 \end{aligned}$$

il polinomio di Stieltjes relativo alla $U(M, R)$:

$$U_r(M, R) = \int_{\sigma'} U(N, R) \phi_r(N, M) d\tau'$$

soddisfa, uniformemente rispetto a r e al variare di M, R nel dominio τ interno a τ' , alle limitazioni

$$\begin{aligned}
 U_r(M, R) &= O(MR^{-(m-2)}), \\
 \frac{\partial U_r(M, R)}{\partial x_i} &= O(MR^{-(m-1)}), \\
 \frac{\partial^2 U_r(M, R)}{\partial x_i \partial x_k} &= O(MR^{-m}).
 \end{aligned}$$

Dimostrazione. a) Poniamo

$$(116) \quad \delta = MR, \quad \rho = NR, \quad t = MN$$

e sia τ'_R il dominio ipersferico di centro R e raggio $\delta/2$, σ'_R l'ipersfera $(R, \delta/2)$, contorno di τ'_R .

Si ricava allora, indicata con K una opportuna costante positiva e con σ_m la misura dell'ipersfera unitaria in S_m ,

$$\begin{aligned}
 (117) \quad |U_r(M, R)| &\leq K r^{m/2} \int_{\tau'} \frac{e^{-r(t^2/d^2)}}{\rho^{m-2}} d\tau' = K \{ r^{m/2} \int_{\tau'_R} \frac{e^{-r(t^2/d^2)}}{\rho^{m-2}} d\tau'_R \\
 &\quad + r^{m/2} \int_{\tau' - \tau'_R} \frac{e^{-r(t^2/d^2)}}{\rho^{m-2}} d\tau' \} < K \{ r^{m/2} e^{-r(\delta^2/4d^2)} \int_{\tau'_R} \frac{d\tau'}{\rho^{m-1}} \\
 &\quad + r^{m/2} (2/\delta)^{m-2} \int_{-\infty}^{\infty} e^{-r((x_1-y_1)^2/d^2)} dy_1 \cdots \int_{-\infty}^{\infty} e^{-r((x_m-y_m)^2/d^2)} dy_m \} \\
 &= K \left\{ r^{m/2} \sigma_m (\delta^2/2) e^{-r(\delta^2/4d^2)} + 2^{m-2} \frac{d^m \pi^{m/2}}{\delta^{m-2}} \right\} = O(\delta^{-(m-2)}),
 \end{aligned}$$

uniformemente rispetto a r e al variare di M, R in τ .

b) Siccome è $|U| \leq K\delta^{-(m-2)}$, con K costante positiva, risulta, supposto che τ' abbia contorno σ' di classe 1,

$$\begin{aligned}
 (118) \quad \frac{\partial U_r(M, R)}{\partial x_i} &= \int_{\tau'} U(N, R) \frac{\partial \phi_r(N, M)}{\partial x_i} d\tau' \\
 &= - \int_{\tau'} U(N, R) \frac{\partial \phi_r(N, M)}{\partial y_i} d\tau' \\
 &= - \int_{\tau'} \frac{\partial}{\partial y_i} (U(N, R) \phi_r(N, M)) d\tau' + \int_{\tau'} \frac{\partial U(N, R)}{\partial y_i} \phi_r(N, M) d\tau' \\
 &= \int_{\sigma'} U(N, R) \phi_r(N, M) \cos(n'y_i) d\sigma' + \int_{\tau'} \frac{\partial U(N, R)}{\partial y_i} \phi_r(N, M) d\tau'
 \end{aligned}$$

dove n' è la normale in N a σ' , orientata verso l'interno di τ' .

Indicata con $h > 0$ la distanza del dominio τ da $\delta\sigma'$, con K' una costante positiva e con μ' la misura di σ' , si ha poi, essendo M ed R punti di τ ,

$$(119) \quad \left| \int_{\sigma'} U(N, R) \phi_r(N, M) \cos(n'y_i) d\sigma' \right| \leq \frac{K'\mu'}{h^{m-2}} r^{m/2} e^{-r(h^2/d^2)} = O(1),$$

$$\begin{aligned}
 (120) \quad \left| \int_{\tau'} \frac{\partial U(N, R)}{\partial y_i} \phi_r(N, M) d\tau' \right| &\leq K' \{ r^{m/2} e^{-r(\delta^2/4d^2)} \int_{\tau'_R} \frac{d\tau'_R}{\rho^{m-1}} \\
 &+ r^{m/2} (2/\delta)^{m-1} \int_{-\infty}^{\infty} e^{-r((x_1-y_1)^2/d^2)} dy_1 \cdots \int_{-\infty}^{\infty} e^{-r((x_m-y_m)^2/d^2)} dy_m \} = O(\delta^{-(m-1)}).
 \end{aligned}$$

Dalle (118), (119), (120) segue allora

$$(121) \quad \frac{\partial U_r(M, R)}{\partial x_i} = O(MR^{-(m-1)}),$$

uniformemente rispetto a r e al variare di M, R in τ .

c) Risulta, per la (118); se n'_R è la normale a σ'_R orientata verso l'esterno di τ'_R ,

$$\begin{aligned}
 (122) \quad \frac{\partial^2 U(M, R)}{\partial x_i \partial x_k} \int_{\sigma'} U(N, R) \frac{\partial \phi_r(N, M)}{\partial x_k} \cos(n'y_i) d\sigma' \\
 - \int_{\tau' - \tau'_R} \frac{\partial U(N, R)}{\partial y_i} \frac{\partial \phi_r(N, M)}{\partial y_k} d\tau' \\
 + \int_{\tau'_R} \frac{\partial U(N, R)}{\partial y_i} \frac{\partial \phi_r(N, M)}{\partial x_k} d\tau'_R \\
 = \int_{\sigma'} U(N, R) \frac{\partial \phi_r(N, M)}{\partial x_k} \cos(n'y_i) d\sigma' \\
 - \int_{\tau' - \tau'_R} \frac{\partial}{\partial y_k} \left(\frac{\partial U(N, R)}{\partial y_i} \phi_r(N, M) \right) d\tau'
 \end{aligned}$$

$$\begin{aligned}
& + \int_{\tau' - \tau'_R} \frac{\partial^2 U(N, R)}{\partial y_i \partial y_k} \phi_r(N, M) d\tau' + \int_{\tau'_R} \frac{\partial U(N, R)}{\partial y_i} \frac{\partial \phi_2(N, M)}{\partial x_k} d\tau'_R \\
= & \int_{\sigma'} \{ U(N, R) \frac{\partial \phi_r(N, M)}{\partial x_k} \cos(n'y_i) \\
& + \frac{\partial U(N, R)}{\partial y_i} \phi_r(N, M) \cos(n'y_k) \} d\sigma' \\
& + \int_{\sigma'_R} \frac{\partial U(N, R)}{\partial y_i} \phi_r(N, M) \cos(n'_R y_k) d\sigma' \\
& + \int_{\tau' - \tau'_R} \frac{\partial^2 U(N, R)}{\partial y_i \partial y_k} \phi_r(N, M) d\tau' \\
& + \int_{\tau'_R} \frac{\partial U(N, R)}{\partial y_i} \frac{\partial \phi_2(N, M)}{\partial x_k} d\tau'_R
\end{aligned}$$

e si ha, indicata con K'' una costante positiva,

$$\begin{aligned}
(123) \quad & \left| \int_{\sigma'} \{ U(N, R) \frac{\partial \phi_r(N, M)}{\partial x_k} \cos(n'y_i) \right. \\
& \left. + \frac{\partial U(N, R)}{\partial y_i} \phi_r(N, M) \cos(n'y_k) \} d\sigma' \right| \\
& \leq K'' \mu' \{ (1/h^{m-2}) r^{m+2/2} e^{-r(\lambda^2/d^2)} + (1/h^{m-1}) r^{m/2} e^{-r(\lambda^2/d^2)} \} = O(1),
\end{aligned}$$

$$\begin{aligned}
(124) \quad & \left| \int_{\sigma'_R} \frac{\partial U(N, R)}{\partial y_i} \phi_r(N, M) \cos(n'_R y_k) d\sigma'_R \right| K'' (r^{m/2}/\delta^{m-1}) e^{-r(\delta^2/4d^2)} \delta^{m-1} \\
& = O(\delta^{-m}),
\end{aligned}$$

$$\begin{aligned}
(125) \quad & \left| \int_{\tau' - \tau'_R} \frac{\partial^2 U(N, R)}{\partial y_i \partial y_k} \phi_r(N, M) d\tau' \right| \leq K'' r^{m/2} (2/\delta)^m \\
& \int_{\tau'} e^{-r(t^2/d^2)} d\tau' = O(\delta^{-m}),
\end{aligned}$$

$$\begin{aligned}
(126) \quad & \left| \int_{\tau'_R} \frac{\partial U(N, R)}{\partial y_i} \frac{\partial \phi_r(N, M)}{\partial x_k} d\tau'_R \right| \leq K'' r^{(m+2/2)} \\
& \int_{\tau'_R} (|x_k - y_k|/\rho^{m-1}) e^{-r(t^2/d^2)} d\tau'_R \\
& < K'' r^{(m+2/2)} e^{-r(\delta^2/4d^2)} 3/2 \delta^2 \sigma_m = O(\delta^{-m}).
\end{aligned}$$

Dalle (122), (123), (124), (125), (126) segue allora, uniformemente rispetto ad r e al variare di M, R in τ ,

$$(127) \quad \frac{\partial^2 U_r(M, R)}{\partial x_i \partial x_k} = O(MR^{-m}).$$

Osservazione. Dalla dimostrazione data in a) si deduce che *se risulta, uniformemente, al variare di M, R in τ' ,*

$$U(M, R) = O(MR^{-k}),$$

con $k < m$, si ha anche, uniformemente rispetto a r e al variare di M, R in τ ,

$$U_r(M, R) = O(MR^{-k}).$$

LEMMA III. Se la funzione $u(M)$ è sommabile in τ , lo è anche la funzione $u(M)MP^{-k}$, con $k < m$, esclusi al più, in τ , i punti P di un insieme di misura nulla.

Dimostrazione. Se τ_p è il dominio ipersferico limitato dall'ipersfera (P, ρ) , si ha quasi ovunque in τ , porto $\rho = PM$,

$$(128) \quad \lim_{\rho \rightarrow 0} 1/\rho^m \int_{\tau_p} |u(M) - u(P)| d\tau_p = 0.$$

Sia P un punto in cui vale la (128). Posto allora

$$\psi(\rho) = \int_{\tau_p} |u(M) - u(P)| d\tau_p,$$

si determini $\delta > 0$ in modo che, per $0 \leq \rho \leq \delta$, risulti $0 \leq \psi(\rho) \leq \rho^m$.

Si ricava allora, essendo σ_m la misura dell'ipersfera unitaria in S_m ,

$$\begin{aligned} \int_{\tau_\delta} (|u(M)|/\rho^k) d\tau_\delta &\leq |u(P)| \int_{\tau_\delta} d\tau_\delta/\rho^k + \int_{\tau_\delta} (|u(M) - u(P)|/\rho^k) d\tau_\delta \\ &= |u(P)| \sigma_m(\delta^{m-k}/m - k) + \int_0^\delta (\psi'(\rho)/\rho^k) d\rho \\ &= |u(P)| \sigma_m(\delta^{m-k}/m - k) + \psi(\delta)/\delta^k + k \int_0^\delta (\psi(\rho)/\rho^{k+1}) d\rho \\ &\leq |u(P)| \sigma_m(\delta^{m-k}/m - k) + \delta^{m-k} + k(\delta^{m-k}/m - k) \end{aligned}$$

da cui la tesi.

2. Indichiamo ora, con il seguente teorema, una notevole proprietà delle funzioni dell'insieme Γ_a .

II. Se w è una funzione di classe 2 nel dominio τ' e u è una funzione dell'insieme Γ_a , vale la formula di Green

$$-\int_{\tau} uE^*(w) d\tau = \int_{\sigma} \{A(\partial w/\partial \nu - Lw) - Bw\} d\sigma - \int_{\tau} f w d\tau.$$

Dimostrazione. La funzione w è in τ' un integrale di classe 2 dell'equazione $E^*(v) = E^*(w)$. Possiamo inoltre supporre che il dominio τ' contenente τ nel suo interno abbia contorno di classe 1 e sia interno a un dominio τ'' nel quale è definita la soluzione fondamentale.

Si ricava allora, per R interno a τ' ,

$$\begin{aligned} k_m w(R) = & \int_{\sigma'} \{w(N) ((\partial F(R, N)/\partial v') + L(N)F(R, N)) \\ & - (\partial w(N)/\partial v')F(R, N)\} d\sigma' \\ & - \int_{\tau'} E^*(w(N))F(R, N) d\tau', \end{aligned}$$

dove v' è la conormale a σ' , orientata verso l'interno di τ' , e N è il punto, di σ' o di τ' , rispetto alle cui coordinate si effettuano le integrazioni.

Ne segue, per le (15), (16) e osservando che i contorni σ e σ' hanno tra loro distanza positiva,

$$\begin{aligned} & \int_{\sigma} \{A(M) (\partial w(M)/\partial v) - (B(M) + L(M)A(M)w(M))\} d\sigma \\ & \quad - \int_{\tau} f(M)m(M) d\tau \\ = & \int_{\sigma} \{A(M) \left[\frac{1}{k_m} \frac{\partial}{\partial v} \int_{\sigma'} (w(N) ((\partial F(M, N)/\partial v') + L(N)F(M, N) \right. \\ & \quad \left. - (\partial w(N)/\partial v')F(M, N)) d\sigma' \right. \\ & \quad \left. - \frac{1}{k_m} \frac{\partial}{\partial v} \int_{\tau'} E^*(w(N))F(M, N) d\tau' \right] \\ & - (B(M) + L(M)A(M)) \left[\frac{1}{k_m} \int_{\sigma'} (w(N) ((\partial F(M, N)/\partial v') \right. \\ & \quad \left. + L(N)F(M, N)) - (\partial w(N)/\partial v')F(M, N)) d\sigma' \right. \\ & \quad \left. - \frac{1}{k_m} \int_{\tau'} E^*(w(N))F(M, N) d\tau' \right] \} d\sigma \\ & - \int_{\tau} f(M) \left[\frac{1}{k_m} \int_{\sigma'} (w(N) ((\partial F(M, N)/\partial v') + L(N)F(M, N) \right. \\ & \quad \left. - (\partial w(N)/\partial v')F(M, N)) d\sigma' \right. \\ & \quad \left. - \frac{1}{k_m} \int_{\tau'} E^*(w(N))F(M, N) d\tau' \right] d\tau \\ = & \int_{\sigma'} \{w(N) (\partial/\partial v' \left[\frac{1}{k_m} \int_{\sigma} (A(M) ((\partial F(M, N)/\partial v) \right. \\ & \quad \left. - L(M)F(M, N) - B(M)F(M, N)) d\sigma \right. \\ & \quad \left. - \frac{1}{k_m} \int_{\tau} f(M)F(M, N) d\tau \right] \\ & - ((\partial w(N)/\partial v') - L(N)w(N)) \left[\frac{1}{k_m} \int_{\sigma} (A(M) ((\partial F(M, N)/\partial v) \right. \\ & \quad \left. - L(M)F(M, N) - B(M)F(M, N)) d\sigma \right. \end{aligned}$$

$$\begin{aligned}
& -1/k_m \int_{\tau} f(M) F(M, N) d\tau] \} d\sigma' \\
& - \int_{\tau'} E^*(w(N)) [1/k_m \int_{\sigma} (A(M) ((\partial F(M, N)/\partial v) \\
& \quad - L(M) F(M, N) - B(M) F(M, N)) d\sigma \\
& - 1/k_m \int_{\tau} f(M) F(M, N) d\tau] d\tau' = - \int_{\tau} u E^*(w) d\tau
\end{aligned}$$

ciò che prova la tesi.

3. Consideriamo in τ' la successione di funzioni

$$(129) \quad w_r = x_1^{\alpha_1} \cdot \dots \cdot x_m^{\alpha_m} \quad (\alpha_i = 0, 1, \dots).$$

Applicando formalmente la formula di Green alle funzioni u , w_r e posto, al solito, su σ , $u = A$, $\partial u/\partial v = B$, ricaviamo il sistema di infinite equazioni di Fischer-Riesz

$$(130) \quad - \int_{\tau} u E^*(w_r) d\tau = \int_{\sigma} \{A((\partial w_r/\partial v) - L w_r) - B w_r\} d\sigma - \int_{\tau} f w_r d\tau$$

per le quali vale il seguente teorema.

III. *Condizione necessaria e sufficiente perchè le funzioni u , A , B sommabili, la prima su τ , le rimanenti su σ , soddisfino al sistema (130) è che la stesse funzioni, alterando al più il valore della u in un insieme di misura nulla, soddisfino alle (15), (16).*

Dimostrazione. Che la condizione sia sufficiente segue dal Teorema II.

Dimostriamo che la condizione è necessaria. Per questo cominciamo con l'osservare che, se u , A , B soddisfano al sistema (130), soddisfano anche all'equazione

$$(131) \quad - \int_{\sigma} u E^*(F_r) d\tau = \int_{\sigma} \{A((\partial F_r/\partial v) - L F_r) - B F_r\} d\sigma - \int_{\tau} f F_r d\tau$$

essendo F_r un qualsiasi polinomio nelle variabili (x_1, \dots, x_m) .

Assumiamo, in particolare, come polinomio F_r il polinomio di Stieltjes relativo alla soluzione fondamentale $F(M, R)$,¹² definito dall'eguaglianza

¹² Se la $F(M, P)$ data dalla (76) non soddisfa, come funzione di P , alla $E(u) = 0$, si può sostituirla con la funzione di Green $G(M, P)$ relativa a un problema al contorno per la (4) in un dominio τ'' contenente τ al suo interno (M. Gevrey, "Détermination et emploi des fonctions de Green dans les problèmes aux limites relatifs aux équations linéaires du type elliptique," *Journal des Mathématiques*, vol. 9 (1930), p. 11).

$$F_r(M, R) = \int_{\tau'} F(N, R) \phi_r(N, M) d\tau',$$

essendo $\phi_r(N, M)$ il polinomio considerato nel Lemma II.

Per noti teoremi di Tonelli,¹³ il polinomio $F_r(M, R)$ converge uniformemente, per $r \rightarrow \infty$, alla $F(M, R)$ in ogni dominio interno a τ' ma con contenente R . In tale dominio anche le derivate $\partial F_r / \partial x_i$, $\partial^2 F_r / \partial x_i \partial x_k$, convergono uniformemente a $\partial F / \partial x_i$, $\partial^2 F / \partial x_i \partial x_k$.

Sia ora R esterno a τ , poniamo cioè $R = Q$.

Risulta allora uniformemente, al variare di M in τ ,

$$\lim_{r \rightarrow \infty} E^*(F_r(M, Q)) = E^*(F(M, Q)) = 0$$

e quindi dalla (131) si deduce, facendo divergere r ,

$$0 \equiv \int_{\sigma} \{A(M) ((\partial F(M, Q) / \partial \nu) - L(M) F(M, Q)) - B(M) F(M, Q)\} d\sigma \\ - \int_{\tau} f(M) F(M, Q) d\tau$$

cioè le funzioni A , B soddisfano alla (15).

Sia ora R interno a τ ; cioè si ponga $R = P$. Per ipotesi, la funzione $u(M)$ è sommabile in τ ; per il Lemma III, lo è anche la funzione $u(M) MP^{-k}$, con $k < m$, per tutti i punti P di τ in cui è verificata la (128), e cioè quasi ovunque in τ .

Sia P uno di tali punti. Risulta, per la (2),

$$(132) \quad E^*(F_r(M, P)) = \sum a_{ik}(M) \frac{\partial^2 F_r(M, P)}{\partial x_i \partial x_k} \\ + \sum b_i^*(M) \frac{\partial F_r(M, P)}{\partial x_i} + c^*(M) F_r(M, P) \\ = \sum a_{ik}(P) \frac{\partial^2 F_r(M, P)}{\partial x_i \partial x_k} \\ + \sum (a_{ik}(M) - a_{ik}(P)) \frac{\partial^2 F_r(M, P)}{\partial x_i \partial x_k} \\ + \sum b_i^*(M) \frac{\partial F_r(M, P)}{\partial x_i} + c^*(M) F_r(M, P)$$

e osserviamo che, essendo

¹³ L. Tonelli, "Sulla rappresentazione analitica della funzioni di più variabili reali," *Rendiconti del Circolo Matematico di Palermo*, vol. 29 (1910), p. 14 e pp. 20-24.

$$\begin{aligned}
 F(M, P) &= O(MP^{-(m-2)}), \\
 \frac{\partial F(M, P)}{\partial x_i} &= O(MP^{-(m-1)}), \\
 \frac{\partial^2 F(M, P)}{\partial x_i \partial x_k} &= O(MP^{-m}),
 \end{aligned}$$

uniformemente al variare di M, P in τ , per il lemma II le stesse limitazioni valgono per $F_r, \partial F_r / \partial x_i, \partial^2 F_r / \partial x_i \partial x_k$, uniformemente rispetto a r ed al variare di M, P in τ .

Dalla (132), indicato con E_0 l'operatore differenziale (a coefficienti indipendenti da x_1, \dots, x_m) $\Sigma a_{ik}(P)(\partial^2 / \partial x_i \partial x_k)$, segue perciò

$$(133) \quad E^*(F_r(M, P)) = E_0(F_r(M, P)) + \chi_r(M, P)$$

e si ha

$$(134) \quad |\chi_r(M, P)| \leq KMP^{-(m-1)}$$

con K costante positiva indipendente da r, M, P .

Siccome $u(M)MP^{-(m-1)}$ è sommabile in τ , si ricava allora dalla (134), per il teorema di Lebesgue sull'integrazione per serie,

$$\begin{aligned}
 (135) \quad \lim_{r \rightarrow \infty} \int_{\tau} u(M) \chi_r(M, P) d\tau &= \int_{\tau} u(M) \{ \Sigma (a_{ik}(M) - a_{ik}(P)) \frac{\partial^2 F(M, P)}{\partial x_i \partial x_k} \\
 &\quad + \Sigma b_{*i}^*(M) \frac{\partial F(M, P)}{\partial x_i} + c^*(M) F(M, P) \} d\tau \\
 &= - \int_{\tau} u(M) E_0(F(M, P)) d\tau,
 \end{aligned}$$

essendo $F(M, P)$ un integrale della (4).

Si ha poi, per $M \neq P$ e indicato con E'_0 l'operatore $\Sigma a_{ik}(P)(\partial^2 / \partial y_i \partial y_k)$,

$$\begin{aligned}
 (136) \quad E_0(F_r(M, P)) &= \Sigma a_{ik}(P)(\partial^2 / \partial x_i \partial x_k) \int_{\tau'} F(N, P) \phi_r(N, M) d\tau' \\
 &= \int_{\tau'} F(N, P) E'_0(\phi_r(N, M)) d\tau'.
 \end{aligned}$$

Sia τ'_R un dominio ipersferico col centro in P e raggio $\epsilon < MP$, σ'_P l'ipersfera (P, ϵ) , ν'_0 la direzione conormale, relativa all'operatore E'_0 , orientata verso l'interno di $\tau' - \tau'_P$.

Si ha, per la formula di Green,

$$\begin{aligned}
 (137) \quad & - \int_{\tau'-\tau'_P} \{F(N, P) E'_0(\phi_r(N, M)) - \phi_r(N, M) E'_0(F(N, P))\} d\tau' \\
 & = \int_{\sigma'} \{F(N, P) \frac{\partial \phi_r(N, M)}{\partial v'_0} - \frac{\partial F(N, P)}{\partial v'_0} \phi_r(N, M)\} d\sigma' \\
 & + \int_{\sigma'_P} \{F(N, P) \frac{\partial \phi_r(N, M)}{\partial v'_0} - \frac{\partial F(N, P)}{\partial v'_0} \phi_r(N, M)\} d\sigma'_P
 \end{aligned}$$

e risulta

$$\begin{aligned}
 (138) \quad \lim_{\epsilon \rightarrow 0} \int_{\sigma'_P} \{F(N, P) \frac{\partial \phi_r(N, M)}{\partial v'_0} - \frac{\partial F(N, P)}{\partial v'_0} \phi_r(N, M)\} d\sigma'_P \\
 = k_m \phi_r(P, M).
 \end{aligned}$$

Siccome è $|E'_0(F(N, P))| \leq K' NP^{-(m-1)}$, con K' costante positiva, segue allora dalle (136), (137), (138)

$$\begin{aligned}
 (139) \quad & E_0(F_r(M, P)) = -k_m \phi_r(P, M) \\
 & + \int_{\sigma'} \{\phi_r(N, M) \frac{\partial F(N, P)}{\partial v'_0} - \frac{\partial \phi_r(N, M)}{\partial v'_0} F(N, P)\} d\sigma' \\
 & + \int_{\tau'} \phi_r(N, M) E'_0(F(N, P)) d\tau'.
 \end{aligned}$$

Ora, in virtù della (128) ed essendo $\phi(P, M) = \phi(M, P)$, risulta

$$(140) \quad \lim_{\tau \rightarrow \infty} \int_{\tau} u(M) \phi_r(P, M) d\tau = u(P).$$

Si ha poi, essendo τ interno a τ' ,

$$(141) \quad \lim_{\tau \rightarrow \infty} \int_{\tau} u(M) \left[\int_{\sigma'} \{\phi_r(N, M) \frac{\partial F(N, P)}{\partial v'_0} - \frac{\partial \phi_r(N, M)}{\partial v'_0} F(N, P)\} d\sigma' \right] d\tau = 0$$

e infine, per l'osservazione relativa al Lemma II, essendo $E'_0(F(M, P)) = O(MP^{-(m-1)})$,

$$\begin{aligned}
 (142) \quad \lim_{\tau \rightarrow \infty} \int_{\tau} u(M) \left[\int_{\sigma'} \phi_r(N, M) E'_0(F(N, P)) d\tau' \right] d\tau \\
 = \int_{\tau} u(M) E_0(F(M, P)) d\tau.
 \end{aligned}$$

Dalle (133), (134), (135), (139), (140), (141), (142), segue allora

$$\lim_{\tau \rightarrow \infty} \int_{\tau} u(M) E^*(F_r(M, P)) d\tau = -k_m u(P)$$

e quindi, per la (131), facendo divergere r e tenendo presente il lemma II,

$$k_m u(P) = \int_{\sigma} \{A(M) ((\partial F(M, P)/\partial v) - L(M)F(M, P)) \\ - B(M)F(M, P)\} d\sigma - \int_{\tau} f(M)F(M, P) d\tau$$

cioè le funzioni u , A , B soddisfano alla (16).

COROLLARIO. La successione $\{E^*(w_r)\}$ è chiusa rispetto alla totalità delle funzioni sommabili in τ .

Infatti se risulta, per ogni r ,

$$\int_{\tau} uE^*(w_r) d\tau = 0$$

si soddisfa al sistema (130) ponendo $A \equiv B \equiv f \equiv 0$. Ne segue, per la (16), $u(P) = 0$ quasi ovunque in τ .

Osservazione. In luogo delle funzioni w_r date dalle (129) si possono porre nella (130) le funzioni di una qualsiasi successione $\{w'_r\}$ tali che w_r e le sue derivate parziali prime e seconde si possano approssimare uniformemente in τ , contorno σ incluso, rispettivamente mediante combinazioni lineari della w' e mediante le derivate parziali prime e seconde di queste combinazioni.

Di tale proprietà gode una larga classe di successioni $\{w'_r\}$ chiuse rispetto alla totalità delle funzioni sommabili in un dominio τ' contenente τ nel suo interno. Ad esempio, si può porre

$$w'_r = e^{2\pi i \sum a_{k,r}(x_k/T_k)}, \quad (\alpha_{k,r} = 0, 1, \dots)$$

essendo $T_k > \delta_k$, massimo valore della differenza $|x_k - x'_k|$ al variare dei punti $M(x_1, \dots, x_m)$, $M'(x'_1, \dots, x'_m)$ in τ .

4. Con le stesse notazioni adoperate nel 4 del I, considerando il problema misto, si ottiene dal sistema (130) il sistema

$$(143) \quad \int_{\tau} uE^*(w_r) d\tau + \int_{\sigma_1} \{A((\partial w_r/\partial v) - w_r(L - (h/k)))\} d\sigma_1 \\ - \int_{\sigma_2} B \left\{ \frac{k}{h} \frac{\partial w_r}{\partial v} + w_r(1 - L(k/h)) \right\} d\sigma_2 \\ = \int_{\sigma_1} (C/k) w_r d\sigma_1 - \int_{\sigma_2} (C/h)((\partial w_r/\partial v) - L w_r) d\sigma_2 \\ + \int_{\tau} f w_r d\tau = c_r$$

dove le costanti c_r sono note.

Definito allora un vettore \mathbf{G} di componenti u in τ , A in σ_1 , B in σ_2 e una successione di vettori $\{\omega_r\}$ di componenti $E^*(w_r)$ in τ , $\{(\partial w_r/\partial v) - w_r(L - (h/k))\}$ in σ_1 , $-\{ \frac{k}{h} \frac{\partial w_r}{\partial v} + w_r(1 - L(k/h)) \}$ in σ_2 , si ricava, per la (143),

$$(144) \quad (\mathbf{G}, \omega_r) = c_r$$

cioè sono noti i coefficienti di Fourier dell'incognito vettore \mathbf{G} rispetto alla successione $\{\omega_r\}$. Viceversa, sia \mathbf{G} una soluzione del sistema (144). Si conoscono allora le funzioni u in τ , A su σ_1 , B su σ_2 e, per le (110), (111) le B , A possono definirsi su tutto σ ; si trova poi, eliminando C dalle (143) mediante le (110), (111), che u , A , B soddisfano al sistema (130) e quindi per il teorema III alle (15), (16). Per il teorema I e per le (110), (111) risulta perciò soddisfatta su σ la (109).

Inoltre, se vale, nell'insieme Γ_a , il teorema di unicità per il problema considerato, la successione $\{\omega_r\}$ è chiusa rispetto alla totalità dei vettori \mathbf{G} di componenti sommabili; se tale teorema non vale, le autosoluzioni sono caratterizzate dall'essere i corrispondenti vettori \mathbf{G} ortogonali alla successione $\{\omega_r\}$. Infine, se il problema non ammette soluzione, nemmeno il sistema (144) ammette soluzione.

Dedotta poi dalla successione $\{\omega_r\}$ una equivalente successione $\{\xi_r\}$ ortogonale e normale, il sistema (144) equivale al sistema

$$(\mathbf{G}, \xi_r) = d_r$$

dove le costanti d_r risultano note. Perciò *condizione necessaria e sufficiente perchè il problema ammetta soluzione, il vettore \mathbf{G} risultando di norma integrale finita, è che converga la serie $\sum d_r^2$.*

Considerazioni del tutto analoghe si possono svolgere per il problema di Cauchy, nel quale il vettore incognito \mathbf{G} ha componenti u in τ , A in σ_2 , B in σ_2 e il vettore ω_r ha componenti $E^*(w_r)$ in τ , $(\partial v_r/\partial v) - Lv_r$ in σ_2 , $-v_r$ in σ_2 .

SUR UNE SUITE DE QUADRIQUES ASSOCIÉE A UNE CONGRUENCE W .*

par LUCIEN GODEAUX.

Dans une note récente, M. Chenkuo Pa¹ a cherché à donner une nouvelle définition de la suite de quadriques que nous avons attachée à tout point non parabolique d'une surface;² il a ensuite indiqué que l'on pouvait de même attacher une suite de quadriques à toute droite d'une congruence W . C'est ce que nous avons fait voici déjà quelques années³ et nous voudrions indiquer brièvement les résultats auxquels nous sommes parvenus.

1. Soit (x) une surface rapportée à ses asymptotiques u, v . Désignons par Q l'hyperquadrique de Klein, appartenant à un espace linéaire S_5 , à cinq dimensions, qui représente les droites de l'espace, par U, V les points de Q qui représentent respectivement les tangentes en un point x de (x) à la ligne u (sur laquelle u varie) et à la ligne v passant par ce point. Tzitzeica⁴ et M. Bompiani⁵ ont montré que les points U, V sont les transformés de Laplace l'un de l'autre. Ils appartiennent donc à une suite de Laplace

$$(1) \quad \dots, U_n, \dots, U_1, U, V, V_1, \dots, V_n, \dots,$$

chaque point étant le transformé du précédent dans le sens des u . Cette suite est autopolaire par rapport à Q ; le point U_n , par exemple, est le pôle de l'hyperplan $V_{n-2}V_{n-1}V_nV_{n+1}V_{n+2}$. Il en résulte que les plans $U_nU_{n+1}U_{n+2}$ et $V_nV_{n+1}V_{n+2}$ sont conjugués par rapport à Q ; ils coupent cette hyperquadrique suivant deux coniques qui représentent les deux séries réglées d'une quadrique Φ_n . Pour $n = 0, 1, 2, \dots$, on obtient la suite de quadriques que nous avons

* Received April 21, 1947.

¹ "A new definition of the Godeaux sequence of quadrics," *American Journal of Mathematics*, vol. 69 (1947), pp. 117-120.

² "Sur les lignes asymptotiques d'une surface et l'espace réglé," *Bulletin de l'Académie royale de Belgique* (1927), pp. 812-826; (1928), pp. 31-41; "La théorie des surfaces et l'espace réglé," *Actualités scientifiques et industrielles*, Paris (1934).

³ "Sur quelques familles de quadriques associées aux points d'une surface," *Annales de la Société Polonaise de Mathématique* (1928), pp. 213-226; "La théorie des surfaces . . .," *loc. cit.*, pp. 21-24.

⁴ *Géométrie projective différentielle des réseaux*, Paris, (1924).

⁵ "Sull'equazione di Laplace," *Rendiconti del Circolo Matematico di Palermo*, vol. 34 (1912), pp. 383-407.

attachée au point x de la surface, la première quadrique étant la quadrique de Lie. Deux quadriques consécutives de la suite se touchent en quatre points qui sont caractéristiques pour les deux quadriques.

La section de Q par l'hyperplan $V_{n-2}V_{n-1}V_nV_{n+1}V_{n+2}$ représente un complexe linéaire et les coordonnées du pôle U_n de l'hyperplan par rapport à Q sont les coefficients de l'équation de ce complexe en coordonnées de droites. On obtient donc une suite de complexes linéaires qui se succèdent dans une suite de Laplace. On peut facilement, au moyen de cette suite de complexes, définir la suite de quadriques $\Phi, \Phi_1, \Phi_2, \dots$, sans passer par l'espace S_3 .

2. Soient j une droite engendrant une congruence $W:(j)$ et $(x), (\bar{x})$ les surfaces focales de cette congruence. Les asymptotiques u, v se correspondent sur les surfaces $(x), (\bar{x})$ et nous attacherons à la surface (x) la suite le Laplace (1), à la surface (\bar{x}) la suite de Laplace analogue

$$(2) \quad \dots, \bar{U}_n, \dots, \bar{U}_1, \bar{U}, \bar{V}, \bar{V}_1, \dots, \bar{V}_n, \dots.$$

Les droites $UV, \bar{U}\bar{V}$ se coupent en un point J , représentant sur Q la droite j de (j) et on sait (Darboux) que le point J décrit un réseau conjugué aux congruences de droites $UV, \bar{U}\bar{V}$. Le point J appartient donc à une suite de Laplace

$$(3) \quad \dots, J_n, \dots, J_1, J, J_{-1}, \dots, J_{-n}, \dots,$$

inscrite dans les suites (1) et (2). D'une manière précise, le point J_n , par exemple, est le point d'intersection des droites $U_{n-1}U_n$ et $\bar{U}_{n-1}\bar{U}_n$.

Les droites $U\bar{U}, V\bar{V}$ se coupent en un point P qui est le pôle par rapport à Q de l'hyperplan $J_2J_1JJ_{-1}J_{-2}$. Le point P appartient à une suite de Laplace

$$(4) \quad \dots, P_n, \dots, P_1, P, P_{-1}, \dots, P_{-n}, \dots,$$

polaire de la suite (3) par rapport à Q . La suite (4) est circonscrite aux suites (1) et (2). Dans les suites (3) et (4), chaque point est le transformé du précédent dans le sens des u et le point P_n est le pôle de l'hyperplan $J_{-n+2}, J_{-n+1}, J_{-n}, J_{-n-1}, J_{-n-2}$.

Les plans $J_{-n-1}J_{-n}J_{-n+1}$ et $P_{n-1}P_nP_{n+1}$ sont conjugués par rapport à Q et coupent cette hyperquadrique suivant deux coniques qui représentent les génératrices des deux modes d'une quadrique Ψ_n . Lorsque n prend toutes les valeurs entières, positives, nulle ou négatives, on obtient une suite de qua-

driques associées à la génératrice j de la congruence (j) . Pour $n = 0$, la quadrique Ψ_0 dégénère en deux plans qui sont les plans focaux de droite j .

Deux quadriques consécutives de la suite

$$(5) \quad \dots, \Psi_{-n}, \dots, \Psi_{-1}, \Psi_0, \Psi_1, \dots, \Psi_n, \dots$$

se touchent en quatre points qui sont des points caractéristiques de ces quadriques.

En remarquant que les coordonnées du point P_n sont les coefficients de l'équation d'un complexe linéaire, on obtient une suite de complexes linéaires se succédant dans une suite de Laplace; cette suite comprend le complexe linéaire osculateur à la congruence (j) le long de la droite j . On pourra en déduire la définition de la suite de quadriques (5) sans passer par l'espace S_6 .

Nous renvoyons à nos notes citées plus haut pour d'autres propriétés.⁶

UNIVERSITÉ DE LIÈGE.

⁶ Voir aussi: Rozet, "Recherches sur les congruences W ," *Mémoires de la Société des Sciences de Liège* (1935), pp. 1-31.

SKEW SETS.*

By R. H. BING.

A skew set is one which is not topologically equivalent to any subset of the surface of a sphere. Kuratowski proved [1] that a compact skew continuous curve which contains only a finite number of simple closed curves contains one of the two following types of curves:

Type 1. A skew curve S_1 is of type 1 provided there exist six distinct points $P_1, P_2, P_3, Q_1, Q_2, Q_3$ such that S_1 is the sum of nine arcs $P_1Q_1, P_1Q_2, \dots, P_3Q_3$ with end points as indicated and with the common part of two of these arcs that intersect each other being an end point of each.

Type 2. A skew curve S_2 is of type 2 provided there exist five distinct points P_1, P_2, P_3, P_4, P_5 such that S_2 is the sum of ten arcs $P_1P_2, P_1P_3, \dots, P_4P_5$ with end points as indicated and with the common part of two of these arcs that intersect each other being an end point of each.

Claytor showed [2] that a compact cyclic continuous curve is skew only if it contains one of these two types of curves. It would be interesting to get a simple characterization of a general planar set. The study of skew sets may lead in that direction.

In this paper we shall consider the following two types of skew sets which are generalizations of the types of skew curves studied by Kuratowski:

Skew set of type 1. The set S_1 is a skew set of type 1 provided there exist six distinct points $P_1, P_2, P_3, Q_1, Q_2, Q_3$ such that S_1 is the sum of nine connected sets $M(P_1Q_1), M(P_1Q_2), \dots, M(P_3Q_3)$ where $M(P_iQ_j)$ contains $P_i + Q_j$ and contains a limit point of $M(P_rQ_s)$ only if $P_i + Q_j$ intersects $P_r + Q_s$.

Skew set of type 2. The set S_2 is a skew set of type 2 provided there exist five points P_1, P_2, P_3, P_4, P_5 such that S_2 is the sum of ten connected sets $M(P_1P_2), M(P_1P_3), \dots, M(P_4P_5)$ where $M(P_iP_j)$ contains $P_i + P_j$ and contains a limit point of $M(P_rP_s)$ only if $P_i + P_j$ intersects $P_r + P_s$.

* Received July 16, 1946; Presented to the American Mathematical Society, August 22, 1946.

The justification for calling these skew sets follows from Theorems 3 and 4.

Because of their lengths, the proofs of the following two theorems are omitted.

THEOREM 1. *Suppose that $P_1, P_2, P_3, Q_1, Q_2, Q_3$ are six distinct points and that $P_1Q_1, P_1Q_2, \dots, P_3Q_3$ are nine arcs with end points as indicated and such that two of these arcs intersect each other only if they have an end point in common. Then the sum of these arcs is not homeomorphic to any plane set.*

THEOREM 2. *Suppose that P_1, P_2, P_3, P_4, P_5 are five distinct points such that $P_1P_2, P_1P_3, \dots, P_4P_5$ are ten arcs with end points as indicated and such that two of these arcs intersect each other only if they have an end point in common. Then the sum of these arcs is not homeomorphic to any plane set.*

The space S considered in the next two theorems has the property that any pair of points of a component of an open set in S can be joined by an arc lying in that open set.

THEOREM 3. *If S contains a skew set of type 1, it contains a skew curve of type 1.*

Proof. Let $M(P_1Q_1), M(P_1Q_2), \dots, M(P_3Q_3)$ be sets in S satisfying the conditions mentioned in the definition of a skew set of type 1. Let $D(P_1Q_1)$ be an open set containing $M(P_1Q_1)$ but no point of $M(P_2Q_2) + M(P_2Q_3) + M(P_3Q_3)$ and let P_1Q_1 be an arc from P_1 to Q_1 in $D(P_1Q_1)$. If we replace $M(P_1Q_1)$ by the arc P_1Q_1 in the sequence $M(P_1Q_1), M(P_1Q_2), \dots, M(P_3Q_3)$, the sum of the resulting sequence is a skew set of type 1. Similarly, we may replace $M(P_1Q_2)$ by an arc P_1Q_2 , $M(P_1Q_3)$ by an arc P_1Q_3 , \dots , and $M(P_3Q_3)$ by an arc P_3Q_3 in such a way that two of the arcs $P_1Q_1, P_1Q_2, \dots, P_3Q_3$ intersect each other only if they have an end point in common.

We shall now show that $P_1Q_1 + P_1Q_2 + \dots + P_3Q_3$ contains nine arcs $P_1O_{11}Q_1, P_1O_{12}Q_2, \dots, P_3O_{33}Q_3$ with end points as indicated such that two of these arcs intersect each other only if they have an end point in common and the common part of two of these arcs that intersect each other has only a finite number of components.

Let P_1Q_1 be $P_1O_{11}Q_1$. We shall determine $P_1O_{12}Q_2$ as follows. Since $P_1O_{11}Q_1 \cdot P_1Q_2$ is a closed and compact set, it can be covered by a finite set of components of $P_1O_{11}Q_1 - P_1O_{11}Q_1 \cdot (P_2Q_1 + P_3Q_1)$. Let $P_1O_{12}Q_2$ be an arc from P_1 to Q_2 in P_1Q_2 plus the sum of these components such that $P_1O_{11}Q_1 \cdot P_1O_{12}Q_2$ has only a finite number of components.

Now $P_1O_{11}Q_1 \cdot P_1Q_3$ can be covered by a finite set T_1 of components of $P_1O_{11}Q_1 - P_1O_{11}Q_1 \cdot (P_2Q_1 + P_3Q_1)$ and $P_1O_{12}Q_2 \cdot P_1Q_3$ can be covered by a finite set T_2 of components of $P_1O_{12}Q_2 - P_1O_{12}Q_2 \cdot (P_2Q_2 + P_3Q_2)$. Let $P_1O_{13}Q_3$ be an arc from P_1 to Q_3 in P_1Q_3 plus the sum of the elements of $T_1 + T_2$ such that each of the sets $P_1O_{11}Q_1 \cdot P_1O_{13}Q_3$ and $P_1O_{12}Q_2 \cdot P_1O_{13}Q_3$ has only a finite number of components.

This process is continued. To get $P_3O_{33}Q_3$, we let T_{ij} be a finite collection of components of $P_iO_{ij}Q_j - P_iO_{ij}Q_j \cdot (P_1O_{11}Q_1 + P_1O_{12}Q_2 + P_2O_{21}Q_1 + P_2O_{22}Q_2)$ covering $P_iO_{ij}Q_j \cdot P_3Q_3$ and let $P_3O_{33}Q_3$ be an arc from P_3 to Q_3 in P_3Q_3 plus the sum of the elements of $T_{13} + T_{23} + T_{31} + T_{32}$ such that the common part of $P_3O_{33}Q_3$ and each of the sets $P_1O_{13}Q_3, P_2O_{23}Q_3, P_3O_{31}Q_1, P_3O_{32}Q_2$ has only a finite number of components. Denote the collection of arcs $P_1O_{11}Q_1, P_1O_{12}Q_2, \dots, P_3O_{33}Q_3$ by G and the sum of the elements of G by G^* . We note that two elements of G intersect each other only if they have an end point in common and the common part of two intersecting elements of G has only a finite number of components.

Some subset of G^* is irreducible with respect to containing six distinct points $P'_1, P'_2, P'_3, Q'_1, Q'_2, Q'_3$ and nine arcs $P'_1Q'_1, P'_1Q'_2, \dots, P'_3Q'_3$ such that two of these arcs intersect each other only if they have an end point in common. For convenience, suppose that G^* is such a subset of itself.

No point of G^* is a point of order five or more. If there were such a point P' , it would belong to some three elements of G and the point P' could be used instead of the common end point of the three arcs of G containing P' . Then G^* would not be irreducible in the sense supposed. Likewise, we find that G^* contains no point of order four that belongs to three elements of G .

Since the common part of no two elements of G has infinitely many components, G^* does not contain infinitely many simple closed curves. We know by Theorem 1 that G^* is a skew set and by the previously mentioned theorem of Kuratowski that it contains either a skew curve of type 1 or a skew curve of type 2.

Assume that G^* contains a skew curve of type 2. Let A_1, A_2, \dots, A_5 be five distinct points and $A_1A_2, A_1A_3, \dots, A_4A_5$ be ten arcs in G^* such that the common part of two of these arcs that intersect each other is an end point of each. If some component of $G^* - (A_1 + A_2 + \dots + A_5)$ has three of

the five points A_1, A_2, \dots, A_5 as limit points, a careful examination of G^* shows that it contains a skew set of type 1. Compare this notion with a theorem by Hall [3]. Assume that no such component contains three of these points as limit points.

Let X be a component of $G - (A_1 + A_2 + \dots + A_5)$ that contains P_1 and has A_i and A_j as limit points. If BA_i and CA_i are arcs in $X + A_i$ each of which is a subset of an element of G , then BA_i and CA_i do not belong to different elements of G or else A_i is a point of order four or more belonging to three elements of G .

There are arcs from P_1 to Q_1 in each of the sets $P_1O_{11}Q_1$, $P_1O_{12}Q_2 + P_2O_{22}Q_2 + P_2O_{21}Q_1$, and $P_1O_{13}Q_3 + P_3O_{33}Q_3 + P_3O_{31}Q_1$ but no two of these sets contains an arc lying in $X + A_i$ and containing A_i ; also no two of these sets contains an arc lying in $X + A_j$ and containing A_j . Therefore Q_1 is a point of X . Likewise, we find that each of the six points $P_1, P_2, P_3, Q_1, Q_2, Q_3$ is a point of X . But $G^* - X$ is not a subset of one arc. Hence, G^* contains a skew curve of type 1.

The following theorem may be established in a like manner.

THEOREM 4. *If S contains a skew set of type 2, it contains either a skew curve of type 1 or a skew curve of type 2.*

Example. That we cannot conclude under the hypotheses of Theorem 4 that S contains a skew curve of type 2 may be seen from the following example showing that a skew curve of type 1 contains a skew set of type 2. Consider the skew curve which is the sum of nine arcs $P_1Q_1, P_1Q_2, \dots, P_3Q_3$ satisfying conditions given in the definition of a skew curve of type 1. Let P_1Q_1 be the sum of the two arcs P_1R and RQ_1 . The sum of the arcs $RQ_1 + P_2Q_1$, $RQ_1 + P_3Q_1$, $P_1R + P_1Q_2$, $P_1R + P_1Q_3$, $P_2Q_1 + P_3Q_1$, P_2Q_2 , P_2Q_3 , P_3Q_2 , P_3Q_3 , and $P_1Q_2 + P_1Q_3$ is a skew set of type 2. However, it contains no skew curve of type 2.

THEOREM 5. *Suppose that $\alpha_1, \alpha_2, \dots, \alpha_n$ are arcs in the plane such that two of these arcs intersect only if they have an end point in common. There exists a collection of arcs $\beta_1, \beta_2, \dots, \beta_n$ in the plane such that β_i ($i = 1, 2, \dots, n$) has the same end points α_i and contains no non end point of β_j ($j \neq i$).*

Proof. Let γ_i ($i = 1, 2, \dots, n$) be an arc (not necessarily in the plane) having the same end points as α_i and such that γ_i contains no non end point

of γ_j ($j \neq i$). We shall show that $\gamma_1 + \gamma_2 + \cdots + \gamma_n$ is topologically equivalent to a plane set. The truth of Theorem 5 will follow as a consequence.

If $\gamma_1 + \gamma_2 + \cdots + \gamma_n$ is not homeomorphic with a plane set, it contains a skew curve of either type 1 or type 2. Assume that it contains nine arcs $P_1Q_1, P_1Q_2, \cdots, P_3Q_3$ satisfying conditions given in the definition of a skew curve of type 1. Then P_iQ_j is the sum of a subcollection of the arcs $\gamma_1, \gamma_2, \cdots, \gamma_n$; suppose that it is the sum of γ_r, \cdots , and γ_s . Let $M(P_iQ_j)$ be the sum of the arcs α_r, \cdots , and α_s . Then the sum of $M(P_1Q_1), M(P_1Q_2), \cdots$, and $M(P_3Q_3)$ is a skew set of type 1. But the plane contains no skew set of type 1. Also, the assumption that $\gamma_1 + \gamma_2 + \cdots + \gamma_n$ contains a skew curve of type 2 leads to the contradiction that $\alpha_1 + \alpha_2 + \cdots + \alpha_n$ contains a skew set of type 2.

Question. It would be interesting to know if the following statement is true: If $\alpha_1, \alpha_2, \cdots, \alpha_n$ are arcs two of which intersect only if they have an end point in common, then there exist arcs $\beta_1, \beta_2, \cdots, \beta_n$ such that two of these arcs intersect only if they have an end point in common, the common part of two of these arcs that intersect is connected, and β_i ($i = 1, 2, \cdots, n$) is an arc in $\alpha_1 + \alpha_2 + \cdots + \alpha_n$ having the same end points as α_i .

THEOREM 6. *No plane set G contains a collection of five mutually separated points sets such that the closure of the sum of any pair of these five sets is the closure of a connected subset of G which is open in G .*

Proof. Assume that the theorem is false. Let S_1, S_2, S_3, S_4, S_5 be five mutually separated sets in the plane set G such that the closure of the sum of any pair of these five sets is the closure of a connected subset of G which is open in G . Denote a point of S_i ($i = 1, 2, 3, 4, 5$) by P_i . Let T_{ij} ($i = 1, 2, 3, 4, 5; j = 1, 2, 3, 4, 5$) be a connected subset of G which is open in G and such that the closure of T_{ij} is the closure of $S_i + S_j$. If $M(P_iP_j)$ is the sum of P_i, P_j , and T_{ij} , the sum of the sequence $M(P_1P_2), M(P_1P_3), \cdots, M(P_4P_5)$ is a skew set of type 2. But the plane contains no skew set of type 2 because it contains no skew curve of either type 1 or type 2.

Problem. The following problem could be posed to a person not acquainted with terms in topology. Consider a nation composed of more than four states. Suppose that there are five political parties, that each state favors some one of these parties, and that each party is favored by some state. Would it be possible for each pair of the parties to hold a joint

convention at some place so that each of the states favoring these two parties could send a delegate to the convention so that this delegate could remain in the states but not cross the boundary of any "unfriendly" state, that is one favoring a party other than one of the two holding the joint convention.

We see by Theorem 6 that the answer is in the negative. We let the sum of the states be G and the sum of the interiors of the states favoring a particular party be one of the five sets.

THE UNIVERSITY OF TEXAS.

REFERENCES

1. C. Kuratowski, "Sur le problème des courbes gauches en topologie," *Fundamenta Mathematicae*, vol. 15 (1930), pp. 271-283.
2. S. Claytor, "Topological immersion of Peanian continua in a spherical surface," *Annals of Mathematics*, vol. 35 (1934), pp. 809-835.
3. D. W. Hall, "A note on primitive skew curves," *Bulletin of the American Mathematical Society*, vol. 49 (1943), pp. 935-936.

FREE LOOPS AND NETS AND THEIR GENERALIZATIONS.*

By GRACE E. BATES.**

INTRODUCTION.

It is the object of this investigation to develop a chapter of loop theory¹ that is parallel to the theory of free Abelian groups and direct sums of Abelian groups, and to the theory of free groups and free products of groups. Like these theories, our study is concerned with two problems: the existence of free loops and of free sums of loops (with and without amalgamated subloops), and the structure of their subloops. In this we succeed fairly completely in so far as we are able to prove the existence of the desired loops in a very comprehensive fashion (even more so than in group theory), and in so far as we are able to prove the theorems analogous to the subgroup theorems of Schreier, Kurosch, and others.

In Section 1, we develop an independent study of free nets (and their generalizations), and this theory, in a somewhat restricted form, is applied in later sections to obtain the basic loop theorems. The greater generality of the net theory seems to indicate the possibility of a wider application of this theory than we have used in this paper. In the study of isotopy, for example, it has been shown² that loops are isotopic (similar) if and only if their associated nets are isomorphic, which suggests that the net approach would be preferable in this study.

In Section 2 we generalize the theorems of Baer and Bol³ to show the equivalence of our theory of half-nets and half-loops. In particular, we prove that this equivalence applies also to homomorphisms in the two theories.

We are able to state in 3, then, as almost immediate consequences of preceding net theorems, the basic theorems on free loops and their generaliza-

* Received June 15, 1940; Presented to the American Mathematical Society, April 26, 1946.

** The author wishes to acknowledge her appreciation of the privilege she had in working on this paper under Professor Reinhold Baer, who suggested the topic and directed its preparation.

¹ For references on loop theory, see Albert (1), (2); Baer (4); Bruck (1), (2); Smiley (1).

The numbers in parentheses refer to the Bibliography at the end of this paper.

² See Baer (3); the term "isotopy" is due to Albert, see Albert (1), (2).

³ See Baer (3) and Bol (1).

tions. In particular, we have the theorem that any half-loop is embeddable in one and essentially only one loop which is freely generated by it. The fundamental theorem of this section is the Subloop Theorem, which has as a corollary the loop analogue to Schreier's theorem. We are able to generalize this theorem to free sums of loops and to obtain a refinement theorem on the decomposition of a loop into free summands.

Throughout Section 3 references are made to corresponding theorems from group theory—a procedure which is later justified in the Appendix. Several interesting contrasts between the two theories are noted. For example, every free loop (except the loop of order one) contains a free subloop of countably infinite rank, whereas a like statement about groups is obviously false for free cyclic groups. Again, although not every semi-group can be embedded in a group, it is true that every semi-loop can be embedded in a loop. Indeed, for any semi-loop which is not itself a loop, the containing loop (generated by the semi-loop) is not at all uniquely determined as is the case for Abelian semi-groups. It is also noteworthy that in many of the loop theorems which have their analogues in group theory, the statements for loops are sharper than those possible for groups—as, for example the theorem on subloops of a free sum of loops and the Generalized Subloop Theorem.

The Appendix contains a general formulation of the concept of “freeness” which comprehends the classical definitions of free group and free Abelian group, etc., as well as the definitions used in this paper for loops.

Notations. If J and K are subsets of the loop L , we denote by $J \cup K$, the set-theoretical sum of J and K ; by $J + K$, the subloop generated by J and K ; and by $J \circ K$, the cross-cut of J and K .

If x is an element of the loop L , and if σ is a homomorphism of L , the image of x under σ is denoted by $x\sigma$; similarly, we denote by $J\sigma$ the image of the subset J of L . We use the notation $L \sim K$, if L is isomorphic to K .

We apply similar notations to nets.

1. NETS AND HALF-NETS.

1.1. Basic concepts and definitions. A *half-net* consists of four different kinds of elements: points, r -lines, s -lines, and t -lines, with the following incidence relations:

- (i) Through every point there passes at most one line of each pencil. (By a *pencil* is meant the set of all lines of any one of the three types, as, for example, the set of all r -lines.)

- (ii) Two lines meet in at most one point.

It is a consequence of (i) that lines in the same pencil do not meet.

In general, we shall use capitals for points and lower case letters for lines; we write $P < h$ for " P lies on h ."

A *net* is a half-net which satisfies the additional requirements:

- (i') Through every point there passes at least one line of each pencil.
 (ii') Two lines of different pencils meet in at least one point.

It is clear that any subset of a net, under the same incidence relations as those prevailing in the net, constitutes a *sub-half-net* of the net.

The above definitions of half-nets and nets permit a set of lines, all of the same pencil, to be considered a half-net (or a net). We exclude, however, these "degenerate nets" from all our considerations.

Definition. If H is a sub-half-net of the half-net K , then H is *closed* in K if the following conditions are satisfied:

- (a) If the point P is in H , and if P lies on h in K , then h is in H .
 (b) If the lines $h \neq h'$ are in H , and if h, h' carry Q in K , then Q is in H .

Clearly, any subnet of a net N is closed in N .

A half-net K is *generated* by its sub-half-net H , if K is the only sub-half-net of K which contains H and is closed in K .

A *homomorphism* ϕ of a half-net M is a single-valued mapping of the elements of M upon the elements of a half-net K , such that points are mapped on points, x -lines are mapped on x -lines, and P on h implies that $P\phi$ is on $h\phi$. It is easily verified that the homomorphic image of a net is a net.

A homomorphism ϕ of a half-net M upon the half-net K is an *isomorphism* if ϕ is 1—1 and if ϕ^{-1} is a homomorphism of K upon M . One-to-oneness is enough to insure that a homomorphism of a net upon another net is an isomorphism, but that this is not the case for proper half-nets is illustrated by the following example: Let M be the half-net consisting of the r -line, h , and the point P not on h , and let K be the half-net consisting of the r -line, k , and the point Q on k . Then M may be mapped 1—1 and homomorphically upon K , but the inverse map is clearly not a homomorphism of K upon M .

THEOREM 1.1. *If ρ and σ are homomorphisms of the half-net M , (into*

a half-net K), if M is generated by its sub-half-net J , and if $j\rho = j\sigma$, for every j in J , then $\rho = \sigma$.

Proof. Let T be the sub-half-net of M consisting of those elements, x , of M for which $x\rho = x\sigma$. Clearly, T contains J . Furthermore, T is closed in M , for: Suppose that P is in T , and $P < h$ in M ; then in K , $P\rho < h\sigma$ and $P\sigma < h\sigma$. But $P\rho = P\sigma$, and in K there is at most one line of the same pencil as h on the point $P\rho = P\sigma$. Hence $h\rho = h\sigma$, and h is in T . Similarly, if $h \neq h'$ are in T , and if h, h' , carry Q in M , then we have $h\rho = h\sigma$ and $h'\rho = h'\sigma$ on $Q\rho$ and on $Q\sigma$ in K , thus implying that $Q\rho = Q\sigma$, and hence Q is in T . Since M is generated by J , and $J \subset T$, then $M = T$; that is, $\rho = \sigma$.

1. 2. Extension chains. A half-net M is an *L-extension* of its sub-half-net J , if:

- (1) Every point of M is in J .
- (2) Any line of M which is not in J lies on at least one point of J .

If M satisfies the additional requirement that every point of M carries at least one line of each pencil, then M is a *complete L-extension* of J .

A half-net K is a *P-extension* of its sub-half-net J , if:

- (1) Every line of K is in J .
- (2) Any point of K which is not in J lies on at least two lines of J .

If K satisfies the additional requirement that every pair of lines of different pencils in K meet in at least one point, then K is a *complete P-extension* of J .

The chain $J = J_0 \subset J_1 \subset \dots \subset J_{2n} \subset J_{2n+1} \subset J_{2n+2} \subset \dots$, where J_{2n+1} is an *L-extension* of J_{2n} , and J_{2n+2} is a *P-extension* of J_{2n+1} , will be called an *extension chain* of J . If the J_i are complete extensions, we call the chain a *maximal extension chain* of J . We use this notation throughout.

It is almost obvious that any sub-half-net J of a net N has a maximal extension chain within N . For, if J_{2n} is constructed in N , then J_{2n+1} may be constructed in N by virtue of the fact that every point of $J_{2n} \subset N$ has one and only one line of each pencil in N ; and if J_{2n+1} is constructed in N , then J_{2n+2} may be constructed in N , since every pair of lines of different pencils in $J_{2n+1} \subset N$ meet in one and only one point in N . Hence, by induction, N contains $J = J_0 \subset J_1 \subset \dots \subset J_i \subset \dots$. Furthermore, it is clear that this maximal extension chain of J within N is uniquely determined by N .

Note that our definition of an extension chain does not require that the J_i be contained in a net, although it will be shown later that, as a consequence of the definition, this is the case.

We denote by $\bigcup J_i = M$, the set-theoretical join of the J_i , where the J_i are terms in an extension chain of $J = J_0$, and we define incidence in M as follows: If P, h , are in M , then $P < h$ in M if there exists an i such that P, h , are in J_i with $P < h$ in J_i . (Note that $P < h$ in J_i implies that $P < h$ in J_{i+k} .) It is clear that, under these incidence relations, $M = \bigcup J_i$ is a half-net. Furthermore, if the J_i form a *maximal* extension chain, then M satisfies the additional requirements for a net, since any element x in M is in some J_i and consequently in all succeeding J_i . Thus, P in M is in some J_{2k+1} in which every point has at least one line of each pencil, and h, h' , of different pencils in M , are both in some J_{2p+2} in which every pair of lines of different pencils meet in at least one point.

Definition. When M , a half-net, can be written in the form: $M = \bigcup J_i$, for some $J = J_0$ contained in M , and with J_{i+1} an L or P -extension of J_i for every i , then we say that J and M are connected by the chain of J_i 's.

LEMMA 1.1. *K is a half-net generated by the half-net J , if and only if, J and K are connected by an extension chain of J .*

Proof. If J and K are connected by an extension chain of J , then $K = \bigcup J_i$, and K is a half-net. We have only to show that K is generated by $J = J_0$. Suppose that there exists a sub-half-net T of K such that $J \subset T \subset K$, and that T is closed in K . Assume that it has already been proved that $J_i \subset T$. If $J_{i+1} = J_i$, then $J_{i+1} = J_i \subset T$; hence suppose that $J_{i+1} \neq J_i$. We distinguish two cases:

(1) $i = 2n$. Let h be a line in J_{2n+1} which is not in J_{2n} . Then h is on at least one point P of J_{2n} . But P is in T , and T is closed in K ; hence h is in T .

(2) $i = 2n + 1$. Let P be a point in J_{2n+2} which is not in J_{2n+1} . Then P is on at least two lines k, k' , of J_{2n+1} . Since k, k' , are in T , and T is closed in K , then P is in T .

By induction, we have $\bigcup J_i = K \subset T \subset K$, and hence $T = K$. That is, K is generated by J .

Conversely, suppose that K is a half-net generated by J ; then we may construct in K an extension chain whose first term is J , as follows:

Let J_{2i+1} contain J_{2i} together with all those lines of K which are on

points of J_{2i} , and let J_{2i+2} contain J_{2i+1} together with all those points of K which are intersections of two or more lines of J_{2i+1} . That is, we might call J_{i+1} a K -complete extension of J_i . Then it is easily verified that the union of the J_i is closed in K , so that we have $K = \bigcup J_i$.

COROLLARY. *N is a net generated by the half-net J , if and only if, N and J may be connected by a maximal extension chain.*

Definition. A *null net* is a net consisting of one point and the three lines, one of each pencil, through this point. A *null half-net* is a half-net consisting of one of the following combinations of elements: (1) one point and no lines; (2) one line and one point on this line; (3) two lines of different pencils and no points; (4) two lines of different pencils and one point which is on these two lines; (5) a null net.

It is easily verified that in any net not a null net, there are at least two lines of each pencil, and every line carries at least two points.

Definition. M , an L -extension of J , is *open* if it satisfies the additional requirement:

(2') Any line of M which is not in J lies on *at most* one point of J . This requirement, together with condition (2) for an L -extension, implies that any line of M which is not in J , lies on one and only one point of J .

K , a P -extension of J is open if it satisfies the additional requirement:

(2') Any point of K which is not in J lies on *at most* two lines of J . This requirement, together with condition (2) for a P -extension, implies that any point of K which is not in J , lies on two and only two lines of J .

The chain, $J = J^{(0)} \subset J^{(1)} \subset \dots \subset J^{(2n)} \subset J^{(2n+1)} \subset J^{(2n+2)} \subset \dots$, where $J^{(2n+1)}$ is an open L -extension of $J^{(2n)}$, and $J^{(2n+2)}$ is an open P -extension of $J^{(2n+1)}$, will be called an *open extension chain* of J . If the $J^{(i)}$ are *complete* extensions, we call the above chain a *maximal open extension chain*. We shall follow this notation throughout.

Given a half-net J , we may construct a maximal open extension chain of which J is the first member, in the following way⁴: If P in J lacks an x -line, we adjoin to J one and only one x -line, $x(P)$, in such a way that $x(P)$ carries P and no other point of J . This we do for every point P in J , and for $x = r, s, t$. It is obvious from the nature of this construction that $J^{(1)}$,

⁴ See Hall (1), p. 236. Our construction is similar to Hall's construction for completing a partial projective plane. However, it will be apparent later that our construction has the advantage of uniqueness in regard to net homomorphisms.

the set of elements in J , together with the lines just adjoined, is a half-net under the incidence relations of J together with those incurred by the method of adding new lines. Furthermore, since no new points are adjoined in $J^{(1)}$, but every point of J has in $J^{(1)}$ one line of each pencil, it is clear that $J^{(1)}$ is a complete L -extension. We then form $J^{(2)}$ from $J^{(1)}$ by adjoining to each pair of lines h, h' , not meeting in $J^{(1)}$, one and only one point Q , requiring Q to lie on h and h' , but on no other line of $J^{(1)}$. Again it is clear that $J^{(2)}$ is a half-net, and that $J^{(2)}$ constitutes a complete P -extension of $J^{(1)}$. This process may be carried out a denumerable number of times, with each step producing a bona-fide half-net. Thus $J^{(2k+1)}$ is formed by adjoining to $J^{(2k)}$ lines h on points P of $J^{(2k)}$ in the manner indicated above, so that in $J^{(2k+1)}$ every point has one (and only one) line of each pencil; and $J^{(2k+2)}$ is formed from $J^{(2k+1)}$ by adjoining points of intersection Q to lines h, h' , of $J^{(2k+1)}$, in the manner indicated above, so that in $J^{(2k+2)}$ every pair of lines of different pencils meet in one and only one point.

Then by the Corollary to Lemma 1.1, if we form $\bigcup J^{(i)}$, we are assured of a net $N = \bigcup J^{(i)}$, which is generated by J . That is, *any half-net may be embedded, in at least one way, in a net.*

For future reference, we state below several properties of maximal open extension chains. We exclude the null-net and null-half-nets from these considerations.

Let $J^{(0)} \subset J^{(1)} \subset \dots \subset J^{(i)} \subset \dots$ be a maximal open extension chain, and let $N = \bigcup J^{(i)}$; then N is a net generated by $J^{(0)}$, from the above discussion. We enumerate the following properties for this chain:

(1°) If $P \neq Q$ are in $J^{(i)}$, for some i , and if P and Q lie on the same line h in N , then P and Q lie on h in $J^{(i)}$; and if h, h', h'' , all of different pencils, are in some $J^{(j)}$, and h, h', h'' , lie on a point P in N , then they lie on P in $J^{(j)}$.

This property of the chain is obvious from the nature of our construction, and is also valid for partial open extension chains.

(2°) $J^{(i)} = J^{(i+1)}$, for some $i > 0$, if and only if, $J^{(0)} = N$.

If $J^{(0)} = N$, it is obvious that $J^{(i)} = J^{(0)}$ for all i , so that we need only to justify the statement in one direction. Suppose that $J^{(i)} = J^{(i+1)}$, for some $i > 0$. Then we have $J^{(2k)} = J^{(2k+1)}$ or $J^{(2k)} = J^{(2k-1)}$, for some $k > 0$. But in $J^{(2k+1)}$ every point carries one and only one line of each pencil, and in $J^{(2k)}$, $k > 0$, every pair of lines of different pencils meet in one and only one point. Hence $J^{(i)} = J^{(i+1)}$ is a net, closed in N , and since

$J^{(0)} \subset J^{(i)} \subset N$, with N generated by $J^{(0)}$, we must have $J^{(i)} = N$. Without loss in generality, suppose that $i = 2k$. Then $J^{(2k-1)} = J^{(2k)} = N$; for, if there were a point in $J^{(2k)}$, but not in $J^{(2k-1)}$, then this point would lie on two and only two lines in $J^{(2k)} = N$, which is impossible, since in a net every point carries three lines. But then $J^{(2k-2)} = J^{(2k-1)} = N$; for, if there were a line x in $J^{(2k-1)}$, but not in $J^{(2k-2)}$, then x would lie on one and only one point in $J^{(2k-1)} = N$, which is impossible, since in a non-null net every line carries at least two points. Since k is a positive integer, we must eventually reach the stage in which $J^{(0)} = J^{(1)} = N$, proving our statement.

As an immediate consequence of property (2°), note that if $J^{(0)} \neq N$, then N has an infinite number of lines and points.

(3°) For i sufficiently large, and $J^{(0)} \neq N$, in $J^{(2i+1)}$ and $J^{(2i+3)}$ lines of each of the three pencils are added.

Since every non-null net has at least two lines of each pencil, then, for some $i > 0$, $J^{(2i)}$ has at least two lines of each pencil. Since $J^{(2i+1)} \neq J^{(2i)}$, there is at least one line adjoined in $J^{(2i+1)}$. Suppose, for definiteness, that this is an r -line, r' ; then in $J^{(2i+1)}$ there must be at least one s -line, say s' , and at least one t -line, say t' , which do not meet r' . In $J^{(2i+2)}$, then, the points $P' = r's'$, and $Q' = r't'$ are adjoined, so that in $J^{(2i+3)}$ the t -line, $t(P')$, and the s -line, $s(Q')$ are adjoined.

(4°) If $J^{(0)} \neq N$, and if p is any positive integer, then there is an integer j , such that the number of points adjoined in $J^{(2j)}$ is greater than p , and the number of lines adjoined in $J^{(2j+1)}$ is greater than p .

For i sufficiently great, $J^{(2i-1)}$ contains at least two lines of each pencil. Suppose that in $J^{(2i)}$, there are k points adjoined; (by (2°), k is necessarily greater than 0). Then in $J^{(2i+1)}$, k lines are adjoined, one on each of the k new points of $J^{(2i)}$. In $J^{(2i+1)}$, for each of these k lines there must be at least two lines, one of each of the remaining two pencils, which do not meet the line just adjoined. Hence in $J^{(2i+2)}$, there must be adjoined at least $2k$ points, and in $J^{(2i+3)}$, at least $2k$ lines. That is, for i sufficiently large, the number of points adjoined in the P -extensions is unbounded, as is the number of lines adjoined in the L -extensions.

(5°) For i sufficiently large, and $J^{(0)} \neq N$, there are points P, Q , with $P \neq Q$, such that P and Q are in $J^{(2i)}$, but not in $J^{(2i-1)}$, satisfying in $J^{(2i)}$ the following conditions: (1) Both P and Q lack an x -line (x , being any given one of the three pencils); (2) P and Q have no common line.

By (4°), for j sufficiently great, there are four different points, R, S, T, U ,

which are in $J^{(2j)}$, but not in $J^{(2j-1)}$. Each of these points is on two and only two lines in $J^{(2j)}$, and hence to each of these points there corresponds one and only one pencil x ($= r, s, t$), such that this point is not on any x -line. Since there exist but three pencils, then at least two points among these four, say R, S , are on lines of the same pencils. We assume, without loss in generality, that neither R nor S is on an r -line. If R and S have no common line, then we have found two points satisfying the requirements of our statement. If R and S lie on the same s -line, then in $J^{(2j+2)}$, the points $P = r(R) \cdot t(S)$ and $Q = t(R) \cdot r(S)$ are adjoined, and clearly, both P and Q lack an s -line, and have no line in common. Similarly, if R and S have, instead, a common t -line, then $P' = r(R) \cdot s(S)$ and $Q' = s(R) \cdot r(S)$ in $J^{(2j+2)}$ satisfy the requirements of our statement.

1.3. Existence of free extensions. We make the following definition:

Definition. A half-net K is *free over its sub-half-net J* , if every homomorphism of J into a net N may be extended to a homomorphism of K into N . If a half-net K is free over and generated by its sub-half-net J , we shall say that K is *freely generated by J* .

LEMMA 1.2. *Given a half-net J , then there is essentially at most one net which is freely generated by J .*

Proof. Suppose that N, M , are nets which are free over and generated by J . Then we may extend the identity homomorphism of $J \subset N$ upon $J \subset M$ to a homomorphism ϕ of N upon $N\phi \subset M$. Since M is generated by J , then $N\phi = M$. But we may also extend the identity map of $J \subset M$ upon $J \subset N$ to a homomorphism ψ of M upon $M\psi \subset N$, and again we have $M\psi = N$. Then $N\phi\psi = N$, and $j\phi\psi = (j\phi)\psi = j\psi = j$, for all j in J . By Theorem 1.1, then $\phi\psi = 1$. Similarly, we have $\psi\phi = 1$. That is, ϕ and ψ are reciprocal isomorphisms of N and M , and we have $N \sim M$.

LEMMA 1.3. *If the half-net R is freely generated by its sub-half-net S , and if the sub-half-net T of R contains S and is generated by S , then: (1) R is freely generated by T , and (2) T is free over S .*

Proof. (1) Given a homomorphism ψ of T , then ψ induces a homomorphism ψ_S of S . Since R is free over S , ψ_S may be extended to a homomorphism ϕ of R , and ϕ induces ϕ_T in T . But we have, then, $s\phi_T = s\phi = s\psi_S = s\psi$, for every s in S . Hence (Theorem 1.1), $\phi_T = \psi$ and ϕ is an extension of ψ to a homomorphism of R ; that is, R is free over T . Clearly, R is generated by T , since R is generated by $S \subset T$.

(2) Let σ be a homomorphism of S ; then σ may be extended to a homomorphism ρ of R , and ρ induces ρ_T in T . We have $s\sigma = s\rho_T$, for every s in S , and hence ρ_T is an extension of σ to a homomorphism of T . That is, T is free over S .

The next two lemmas are preparatory to the proof of the existence of a net which is freely generated by a half-net.

LEMMA 1.4. (a). *K , an L -extension of the half-net J , is free over J , if and only if, K is an open L -extension.*

Proof. We assume, first, that K is an open L -extension.

Given η , a homomorphism of J into a net N , we define a mapping ϕ of K into N as follows: Let ϕ act on elements of J in the same manner as η . Consider an element x which is in K , but not in J ; then x is a line in K , lying on one and only one point P of J , since K is an open extension. In N , $P\phi = P\eta$ has one and only one line h of the same pencil as x ; we define $x\phi = h$. That ϕ , as so defined, is single-valued is a consequence of the fact that x lies on P and on no other point in K . Clearly, then, ϕ is a homomorphism of K into N , and ϕ is an extension of η . Hence K is free over J .

Now assume that K is free over J .

Suppose that K is not an open extension of J . Then there must be at least one line h in K , but not in J , such that h lies on P, Q , of J , with $P \neq Q$. But now consider the net $N = \bigcup J^{(i)}$, where $J^{(0)} = J$, and where $J^{(i+1)}$ is a complete open extension (L or P), of $J^{(i)}$. Take the identity map of $J \subset K$ upon $J \subset N$. Since K is free over J , this map may be extended to a homomorphism ψ of K into N . In N , $h\psi$ carries $P\psi = P$, and $Q\psi = Q$. But, since P and Q are both in J , $P \neq Q$, then $h\psi$ is in J on P and Q (Property (1°) for maximal open extension chains). Hence P (and Q) in J had already one line $h\psi$ of the same pencil as h . Then $h = h\psi$, and h is in J , which is a contradiction to our hypothesis. Hence K is open.

LEMMA 1.4. (b). *M , a P -extension of J , is free over J , if and only if, M is an open P -extension of J .*

Proof. Assume that M is an open P -extension of J .

Given η , a homomorphism of J into a net N , we define a mapping ϕ of M into N as follows: Let ϕ act on elements of J in the same manner as η . Consider an element x which is in M , but not in J ; then x is a point of M , lying on two and only two lines, h, h' , of different pencils in J . The lines $h\phi = h\eta$ and $h'\phi = h'\eta$ in N have one and only one point of intersection, Q .

Define $x\phi = Q$. That ϕ as so defined is single-valued, follows from the fact that x lies on h, h' , and on no other line in M . Then, clearly, ϕ is a homomorphism of M into N , and ϕ coincides with η in J ; hence M is free over J .

Now assume that M is free over J .

Suppose that M is not an open P -extension. Then there must be at least one point P in M , but not in J , which lies on three different lines, h, h', h'' , of J . Consider the net $N = \bigcup J^{(i)}$, where $J^{(0)} = J$, and where $J^{(i+1)}$ is a complete open (L or P) extension of $J^{(i)}$. The identity map of $J \subset M$ upon $J \subset N$ may be extended to a homomorphism ψ of M into N , since M is free over J . In N , $P\psi$ lies on $h\psi, h'\psi, h''\psi$,—that is, on h, h', h'' ; since h, h', h'' , are all in J , then $P\psi$ is in J (Property (1°) for maximal open extension chains). Hence h, h', h'' , intersect already in one point $P\psi$ of $J \subset M$. Then $P\psi = P$, and P is in J , which is a contradiction to our hypothesis. Hence M is open.

LEMMA 1.5. *N is a half-net which is freely generated by the half-net J , if and only if, N may be represented in the form: $N = \bigcup J^{(i)}$, where $J^{(0)} = J$, and where $J^{(i+1)}$ is an open (L or P) extension of $J^{(i)}$.*

Proof. Assume, first, that we have $N = \bigcup J^{(i)}$; by previous arguments, then N is a half-net which is generated by $J = J^{(0)}$. Hence, we have only to show that N is free over J .

Given η , a homomorphism of J into a net M , suppose that η has already been extended to a homomorphism η_{2n} of $J^{(2n)}$ into M . By Lemma 1.4 (a), then η_{2n} may be extended to a homomorphism η_{2n+1} of $J^{(2n+1)}$ into M , and by Lemma 1.4. (b), η_{2n+1} may be extended to a homomorphism η_{2n+2} of $J^{(2n+2)}$ into M . Hence, by induction, there exists, for every i , a homomorphism η_i of $J^{(i)}$, such that η_i and η_{i+k} coincide in $J^{(i)}$. We define a mapping ϕ of N into M as follows: If x is in N , then x is in $J^{(i)}$ for some i , and we define $x\phi = x\eta_i$. Then ϕ is single-valued, since for x in $J^{(i)}$ and in $J^{(j)}$, $i < j$, we have $x\eta_i = x\eta_j = x\phi$. Furthermore, ϕ is a homomorphism of N , since $P < h$ in N implies that $P < h$ in some $J^{(i)}$, and hence $P\eta_i = P\phi$ lies on $h\eta_i = h\phi$. Hence, since η has been extended to a homomorphism of N , N is free over J .

Conversely, suppose that there is given a half-net N which is freely generated by its sub-half-net J . By Lemma 1.1, then N may be represented in the form $N = \bigcup J_i$, with $J = J_0$ and where J_{i+1} is an L or P -extension of J_i . We wish to show that these are open extensions.

For any J_i in the chain, we have $J \subset J_i \subset N$, and by hypothesis N is free over J . Since J_i is generated by J , J_i is free over J (Lemma 1.3, (2)).

But then, for every i , we have $J \subset J_i \subset J_{i+1}$, where J_{i+1} is free over J_i (Lemma 1.3,(1)). By Lemmas 1.4,(a) and (b), then, the J_i are open extensions; that is, in our notation, we have $J_i = J^{(i)}$ and $N = \bigcup J^{(i)}$.

COROLLARY. *N is a net which is freely generated by the half-net J , if and only if, N may be represented in the form: $N = \bigcup J^{(i)}$, where $J^{(0)} = J$, and $J^{(i+1)}$ is a complete open (L or P) extension of $J^{(i)}$.*

This corollary is an immediate consequence of Lemma 1.5 and the corollary to Lemma 1.1.

THEOREM 1.2. *Any half-net J may be embedded in one and essentially only one net N which is freely generated by J .*

We have already noted that any half-net J may be embedded in a net N , where N may be represented in the form: $N = \bigcup J^{(i)}$ (see the construction following the Corollary to Lemma 1.1). Then Theorem 1.2 is an immediate consequence of the corollary to Lemma 1.5 and Lemma 1.2.

THEOREM 1.3. *Given a half-net J which is not a net and not a null-half-net, then there exists an extension of J to a net M which is generated by J , but not free over J (Thus, there exist non-equivalent extensions of (proper) half-nets to nets).*

Proof. Form the maximal open extension chain of J :

$$J = J^{(0)} \subset J^{(1)} \subset \dots \subset J^{(i)} \quad \dots,$$

thus obtaining the net $N = \bigcup J^{(i)}$, which is freely generated by J . Since $J \neq N$, by property (5°) for maximal open extension chains there exist, for i sufficiently large, points $P \neq Q$, such that P and Q are in $J^{(2i)}$, but not in $J^{(2i-1)}$, where P, Q , satisfy conditions: (1) Both P and Q lack an x -line, (2) P and Q have no common line. For convenience, suppose that P and Q lack an r -line.

Then we may form the half-net J^* , which is obtained from $J^{(i)}$ by adjoining one r -line connecting P and Q . Clearly, J^* is not free over $J^{(2i)}$ (Lemma 1.4,(b)), and, if we extend J^* to the net M which is freely generated by J^* , then M is generated by J , but is not free over J (Corollary to Lemma 1.5).

1.4. Free nets and free sums of nets. If the net N is freely generated by its sub-half-net J , and if J contains lines, but no points, then we define J as a *free set of generators* of N . A *free net*, then, is a net having a free set of generators.

From Theorem 1.2, it is clear that, given any set of lines J (not all of the same pencil), there is one and essentially only one net N for which J is a free set of generators. We note also, since there are no incidence relations in a free set of generators J of a net N , that any single-valued mapping (which, of course, maps x -lines on x -lines), of J into a net may be extended to a homomorphism of N into that same net.

THEOREM 1.4. *Any net is the homomorphic image of a free net.*

Proof. Consider any net M ; let x be any line in M , belonging, say, to the pencil of s -lines. Then x , together with the set of all points of M which lie on x , form a sub-half-net K of M which generates M . For, every point of M not on x is connected by lines from different pencils with points on x .

Let J be the set consisting of the line x , and all the r -lines through points of x (i. e. J is the set of all r -lines in M , together with the s -line x). Extend J to the free net S having J as a free set of generators. If $J \subset S$ is mapped by the identity map upon $J \subset M$, this map may be extended to a homomorphism ϕ of S into M , and since J generates M , then $M = S\phi$.

Definition. A net N is the *free sum of its subnets* $N(i)$, finite or infinite in number, if the following conditions are satisfied:

(1) N is generated by $\bigcup N(i)$.

(2) If, for every i , there is a homomorphism $\phi(i)$ of $N(i)$ into the net M , then there exists a homomorphism ϕ of N into M , which coincides with $\phi(i)$ in $N(i)$.

We use the notation $N = \sum^* N(i)$ to represent the free sum of the $N(i)$. We write $N(1) * N(2)$ to designate the free sum of $N(1)$ and $N(2)$, and similarly for any finite number of free summands.

An immediate consequence of this definition is that $N(i) \circ N(j)$ is vacuous for all $i \neq j$, where the $N(i)$ are free summands. For, if $N(i)$ and $N(j)$ have a common point, then we map every point P in $N(i)$ upon one and the same point $P\phi(i)$ in M , and all the x -lines in $N(i)$ upon the x -lines through $P\phi(i)$ in M ; and at the same time, we map every point Q in $N(j)$, for all $j \neq i$, upon one and the same point $Q\phi(j) \neq P\phi(i)$, and all the x -lines in these $N(j)$ upon the x -lines on $Q\phi(j)$ in M . It is obvious that $\phi(i)$ and the $\phi(j)$ cannot be extended simultaneously to a homomorphism of $N = \sum^* N(i)$.

Remark. If $N = \sum_{i \neq j}^* N(i)$, then $N = N(i) * M$, where M is the net generated by $\bigcup_{j \neq i} N(j)$, (Theorem 1.1).

If a net M is the free sum of nets $M(i)$, where each $M(i)$ is a free net, then M is a free net. For, each $M(i)$ has a free set of generators, $J(i)$, and it is easily verified that $J = \bigcup J(i)$ is a free set of generators of M .

1.5. Structure of subnets of free nets. We are able to determine, (essentially), completely the structure of subnets of free nets (and their generalizations) by means of the following theorem:

THEOREM 1.5. *If the net N is freely generated by its sub-half-net J , and if S is a subnet of N , then S is the free sum of the nets F and H , where:*

- i. H is freely generated by $S \circ J$.
- ii. F is a free net with a free set of generators:

$$F_0 = \bigcup_{i=0}^{\infty} \{(S \circ J^{(2i+1)}) - (S \circ J^{(2i)})_L\}, \text{ where}$$

$J = J^{(0)} \subset J^{(1)} \subset \dots \subset J^{(2i)} \subset J^{(2i+1)} \subset J^{(2i+2)} \subset \dots$ is the maximal (open) extension chain of J in N .

The proof requires several lemmas. Throughout these discussions we use the subscript, " L ," to denote a complete L -extension within N of a sub-half-net of N , and the subscript, " P ," to denote a complete P -extension.

LEMMA (a). *If $J_0 \subset J_1 \subset \dots \subset J_i \subset \dots$ is a maximal extension chain of J_0 , and if S is a subnet of the net $N = \bigcup_{j=0}^{\infty} J_j$, then:*

$$(1) \quad (S \circ J_{2n})_L \subset S \circ J_{2n+1} \quad (\text{for all } n).$$

$$(2) \quad (S \circ J_{2n+1})_P = S \circ J_{2n+2}$$

Proof. (1) Clearly $S \circ J_{2n} \subset S \circ J_{2n+1}$. Consider an element x , which is in $(S \circ J_{2n})_L$, but not in $S \circ J_{2n}$; then x is a line in J_{2n+1} on at least one point P of $S \circ J_{2n}$. Since P is in S , a subnet of N , then x is also in S ; hence x is in $S \circ J_{2n+1}$, proving (1).

(2) Clearly $S \circ J_{2n+1} \subset S \circ J_{2n+2}$. Consider an element y in $(S \circ J_{2n+1})_P$, but not in $S \circ J_{2n+1}$; then y is a point in J_{2n+2} on at least two lines h, h' , of $S \circ J_{2n+1}$. Since h, h' , are both in the subnet S , then y is in S . Hence y is in $S \circ J_{2n+2}$, and we have $(S \circ J_{2n+1})_P \subset S \circ J_{2n+2}$. If x is an element in

$S \circ J_{2i+2}$, then, if x is also in J_{2n+1} , x is in $S \circ J_{2n+1} \subset (S \circ J_{2n+1})_P$; hence, suppose that x is not in J_{2n+1} . Then x is a point in J_{2n+2} on at least two lines h, h' , of J_{2n+1} . But, since x is in the subnet S , h and h' are in S , so that we have h, h' , in $S \circ J_{2n+1}$. Then x is in $(S \circ J_{2n+1})_P$ and thus $S \circ J_{2n+2} \subset (S \circ J_{2n+1})_P$. Hence, $(S \circ J_{2n+1})_P = S \circ J_{2n+2}$.

COROLLARY. *Under the hypotheses of the Lemma, if J_{2n+1} is an open L -extension of J_{2n} , then $(S \circ J_{2n})_L$ is an open L -extension of $S \circ J_{2n}$; and if J_{2n+2} is an open P -extension of J_{2n+1} , then $S \circ J_{2n+2}$ is an open P -extension of $S \circ J_{2n+1}$.*

This is an immediate consequence of the fact that an element in $(S \circ J_{2n})_L$, but not in $S \circ J_{2n}$, is in J_{2n+1} but not in J_{2n} ; and an element in $S \circ J_{2n+2}$ but not in $S \circ J_{2n+1}$ is in J_{2n+2} but not in J_{2n+1} .

If K and M are sub-half-nets of a net N , such that $K \subset M$, then we denote by $M - K$, the half-net consisting of elements which are in M but not in K , with incidence relations for these elements as prescribed by N .

We say that two sub-half-nets J, K , of a net N are *totally disconnected* if $J \circ K$ is vacuous, and if we have:

- (i) If P is in J , h in K , then $P \not\prec h$ (in N).
- (ii) If k is in J , Q in K , then $k \not\succ Q$ (in N).

Hence, if we have a set of half-nets $J(n)$, which are pairwise totally disconnected in N , then $\bigcup J(n)$ is a sub-half-net of N , which has the property that $P < h$ in $\bigcup J(n)$, if and only if, $P < h$ in $J(n)$, for some n .

LEMMA (b). *If $J_0 \subset J_1 \subset \dots \subset J_i \subset \dots$ is a maximal extension chain generating the net $N = \bigcup J_i$, and if S is a subnet of N , then*

$$K(i) = \{(S \circ J_{2i+1}) - (S \circ J_{2i})_L\}, i = 0, 1, 2, \dots$$

has the following properties:

- (1) $K(i)$ is a set of lines from S , and contains no points.
- (2) $K(i) \circ K(j)$ is vacuous for $i \neq j$.
- (3) $(S \circ J_{2n})_L$ and $\bigcup_{i \geq n} K(i)$ are totally disconnected.

Proof. (1) By Lemma (a), $(S \circ J_{2i})_L \subset S \circ J_{2i+1}$, so that the expressions for $K(i)$ have meaning. Clearly, $K(i) \subset S$. If x is in $K(i)$ then x is in $S \circ J_{2i+1}$, but not in $(S \circ J_{2i})_L$. But then x is in J_{2i+1} and not in J_{2i} , so that x is a line.

(2) Since $K(i) \subset J_{2i+1} - J_{2i}$, it is obvious that $K(i) \cap K(j)$ must be vacuous for $i \neq j$.

(3) The set $(S \cap J_{2n})_L$ is excluded from $K(n)$; since we have $(S \cap J_{2n})_L \subset S \cap J_{2n+1}$, it is also excluded from all the $K(i)$, for $i > n$. Hence $(S \cap J_{2n})_L \cap \bigcup_{i \geq n} K(i)$ is vacuous. Now consider a point P in $(S \cap J_{2n})_L$; since $(S \cap J_{2n})_L$ is a complete L -extension of $S \cap J_{2n}$ then $r(P)$, $s(P)$, $t(P)$, are all in $(S \cap J_{2n})_L$, and hence not in $\bigcup_{i \geq n} K(i)$. That is, for P in $(S \cap J_{2n})_L$, there is no h in $\bigcup_{i \geq n} K(i)$, such that $P < h$. Since $\bigcup_{i \geq n} K(i)$ contains no points, there is no Q in $\bigcup_{i \geq n} K(i)$, such that $Q < k$, for k in $(S \cap J_{2n})_L$. Hence we have proved (3).

LEMMA (c). If $J_0 \subset J_1 \subset \dots \subset J_i \subset \dots$ is a maximal extension chain generating the net $N = \bigcup J_i$; and if S is a subnet of N , then S is generated by

$$M = (S \cap J_0) \cap \bigcup_{i \geq 0} \{(S \cap J_{2i+1}) - (S \cap J_{2i})_L\}.$$

Proof. Define:

$$M(2n) = (S \cap J_{2n}) \cap \bigcup_{i \geq n} K(i); \quad M(2n+1) = (S \cap J_{2n+1}) \cap \bigcup_{i \geq n+1} K(i),$$

where, as in Lemma (b), $K(i) = \{(S \cap J_{2i+1}) - (S \cap J_{2i})_L\}$. Thus $M = M(0)$.

S may be represented in the form: $S = \bigcup (S \cap J_i)$ —for, clearly $\bigcup (S \cap J_i) \subset S$, and S is also contained in $\bigcup (S \cap J_i)$, since x in $S \subset N = \bigcup J_i$ is, for some n , in $S \cap J_n \subset \bigcup (S \cap J_i)$.

Now consider $\bigcup M(i)$; since each $M(i)$ is a sub-half-net of S , we have $\bigcup M(i) \subset S$. But, since $S = \bigcup (S \cap J_i)$, then $S \subset \bigcup M(i)$, and we have S also represented in the form: $S = \bigcup M(i)$.

Suppose that there is a net R , such that $M \subset R \subset S = \bigcup M(i)$. Assume that it has already been shown that $M(n)$ is contained in R ; we distinguish two cases:

$$(1) \quad n = 2k. \quad M(2k) = (S \cap J_{2k}) \cap \bigcup_{i \geq k} K(i) \subset R$$

$$\begin{aligned} M(2k+1) &= (S \cap J_{2k+1}) \cap \bigcup_{i \geq k+1} K(i) \\ &= (S \cap J_{2k})_L \cap \{(S \cap J_{2k+1}) - (S \cap J_{2k})_L\} \cap \bigcup_{i \geq k+1} K(i) \\ &= (S \cap J_{2k})_L \cup \bigcup_{i \geq k} K(i). \end{aligned}$$

Then $M(2k+1)$ contains, in addition to elements of $M(2k)$, only lines which are on at least one point of $(S \cap J_{2k}) \subset M(2k) \subset R$. Since R is a subnet of N , clearly these lines must be in R . That is, $M(2k+1) \subset R$.

$$(2) \quad n = 2k + 1. \quad M(2k + 1) = (S \circ J_{2k+1}) \cup \bigcup_{i \geq k+1} K(i) \subset R.$$

$$M(2k + 2) = (S \circ J_{2k+2}) \cup \bigcup_{i \geq k+1} K(i).$$

Since $(S \circ J_{2k+2}) = (S \circ J_{2k+1})_P$ (by Lemma (a)), $M(2k + 2)$ contains, in addition to elements of $M(2k + 1)$, only points which are on at least two lines of $(S \circ J_{2k+1}) \subset M(2k + 1) \subset R$. Since R is a subnet of N , clearly these points must be in R . That is, $M(2k + 2) \subset R$.

By induction, we have $\bigcup_{i \geq 0} M(i) = S \subset R \subset S$, and hence $R = S$; that is, S is generated by M .

Note that in the chain of $M(i)$'s, $M(2n + 2)$ is not a complete P -extension of $M(2n + 1)$, since the only points added in $M(2n + 2)$ are those lying on two or more lines of $S \circ J_{2n+1}$, which is only a part of $M(2n + 1)$; that is, in $M(2n + 2)$, not every two lines of different pencils may meet. However, from the Corollary to Lemma 1.1, we have that a net is generated by its sub-half-net, if and only if, it is the join of half-nets in a maximal extension chain of which the sub-half-net is the first member. Hence S may be written in the form: $S = \bigcup_{i \geq 0} M_i$, where $M_0 = M$, and where M_{i+1} is a complete (L or P) extension of M_i .

We are now ready to prove Theorem 1.5.

Since N is freely generated by its sub-half-net J , then N may be represented as $\bigcup J^{(i)}$, where $J = J^{(0)}$, and, as is our usual notation, $J^{(i+1)}$ is a complete open (L or P) extension of $J^{(i)}$ (Corollary to Lemma 1.5.). By Lemma (c), the subnet S of N is generated by

$$M = (S \circ J^{(0)}) \cup \bigcup_{i \geq 0} K(i), \text{ where } K(i) = \{(S \circ J^{(2i+1)}) - (S \cup J^{(2i)})_L\}.$$

To show that S is free over M , we consider again the chain of $M(i)$'s, where, as before,

$$M(2n) = (S \circ J^{(2n)}) \cup \bigcup_{i \geq n} K(i); \quad M(2n + 1) = (S \circ J^{(2n+1)}) \cup \bigcup_{i \geq n+1} K(i).$$

By the Corollary to Lemma (a), we have that $(S \circ J^{(2n)})_L$ is an open L -extension of $(S \circ J^{(2n)})$. Since $(S \circ J^{(2n)})_L$ and $\bigcup_{i \geq n} K(i)$ are totally disconnected (Lemma (b)), it is easily verified then, that

$$M(2n + 1) = (S \circ J^{(2n)})_L \cup \bigcup_{i \geq n} K(i)$$

is an open L -extension of $M(2n)$.

Similarly, we consider

$$M(2n + 1) = (S \circ J^{(2n+1)}) \cup \bigcup_{i \geq n+1} K(i),$$

and

$$M(2n+2) = (S \circ J^{(2n+2)}) \vee \bigcup_{i \geq n+1} K(i).$$

By the Corollary to Lemma (a), then $(S \circ J^{(2n+2)})$ is an open P -extension of $(S \circ J^{(2n+1)})$, and, since $(S \circ J^{(2n+2)})$ is contained in $(S \circ J^{(2n+2)})_L$, we know that $(S \circ J^{(2n+2)})$ and $\bigcup_{i \geq n+1} K(i)$ are totally disconnected (Lemma (b)). Hence, as is easily verified, $M(2n+2)$ is an open P -extension of $M(2n+1)$.

Then, in the chain of $M(i)$'s, each $M(i+1)$ is an open extension of $M(i)$, and hence $S = \bigcup_{i \geq 0} M(i)$ is free over M (Lemma 1.5).

Now let H be the subnet of S which is generated by $H_0 = S \circ J$, and let F be the subnet of S which is generated by F_0 , where

$$F_0 = \bigcup_{i=0}^{\infty} \{S \circ J^{(2i+1)} - (S \circ J^{(2i)})_L\}.$$

Let σ be a homomorphism of H , and τ a homomorphism of F into the same net R . Then σ induces a homomorphism σ_0 of H_0 and τ induces a homomorphism τ_0 of F_0 into R . Since H_0 and F_0 are totally disconnected (Lemma (b)), σ_0 and τ_0 define a homomorphism ϕ_0 of $H_0 \vee F_0$ into R , such that ϕ_0 and σ_0 coincide on H_0 , and ϕ_0 and τ_0 coincide on F_0 .

But S is free over and generated by $H_0 \vee F_0 = M$, and hence ϕ_0 may be extended to a homomorphism ϕ of S into R . Then ϕ induces a homomorphism ϕ_H of H ; since ϕ_H and σ coincide on elements of H_0 , and H is generated by H_0 , then $\phi_H = \sigma$ (Theorem 1.1). Similarly, ϕ induces a homomorphism ϕ_F of F , and $\phi_F = \tau$. Hence ϕ is an extension of σ and τ . Furthermore, since $H_0 \vee F_0 \subset H \vee F \subset S$, and S is generated by $H_0 \vee F_0$, clearly S is generated by $H \vee F$. Hence we have shown that $S = H * F$.

We have then only to verify that H is free over $H_0 = S \circ J$ and that F is a free net having a free set of generators F_0 . Let σ_0 be any homomorphism of H_0 into a net R . Then there exists at least one homomorphism τ_0 of F_0 into the same net R (for example, the homomorphism which maps F_0 into a null-net of R). Since S is freely generated by $H_0 \vee F_0 = M$, then the homomorphisms σ_0 and τ_0 may be extended simultaneously to a homomorphism ϕ of S into R . But ϕ induces a homomorphism ϕ_H of H into R and clearly ϕ_H is an extension of σ_0 , which proves that H is free over H_0 . The above argument is symmetric with respect to F and H , so that F is free over F_0 . Then F is clearly a free net, since it has a free set of generators F_0 (by Lemma (b) F_0 consists of lines only).

COROLLARY. *A subnet of a free net is free.*

For, if N is a free net, then N has a free set of generators J , and by the Theorem, a subnet S is of the form $S = H * F$, where F is a free net, and where H is free over and generated by $S \circ J$. Since J is a set of lines, then $S \circ J$ consists entirely of lines, and hence H is a free net. But the free sum of free nets is free, so that S is a free net.

THEOREM 1.6. *N is a free net, if and only if, for every net M and homomorphism ϕ such that $M\phi = N$, there is a subnet R of M , such that ϕ induces an isomorphism of R upon N .*

Proof. Suppose that N is a free net; then N has a free set of generators J , consisting of lines only.

Let M be a net, ϕ a homomorphism, such that $M\phi = N$. Consider any line x in J ; then there is at least one line y in M , such that $x = y\phi$. For each x in J , select one and only one y in M , with $y\phi = x$. Let R_0 be the sub-half-net of M consisting of these lines y .

Then, clearly, ϕ induces a homomorphism ϕ_0 of R_0 upon J which is 1 — 1. Since J has no incidence relations, ϕ_0^{-1} is a homomorphism of J upon R_0 , and hence ϕ_0 is an isomorphism of R_0 upon J .

Since N is free over J , ϕ_0^{-1} may be extended to a homomorphism μ of N upon $N\mu \subset M$; let $N\mu = R$. But ϕ induces a homomorphism ψ of R upon $R\psi$, and we have $R_0\psi = J \subset R\psi \subset N$; since J generates N , then $R\psi = N$.

Hence $N\mu\psi = N$, and $j\mu\psi = (j\mu)\psi = (j\phi_0^{-1})\psi = j\phi_0^{-1}\phi_0 = j$, for every $j \in J$. Then $\mu\psi = 1$ (Theorem 1.1). But, since every y in $R = N\mu$ is of the form $x\mu$ for some x in N , we have that, for all y in R , $y\psi\mu = (x\mu)\psi\mu = (x\mu\psi)\mu = x\mu = y$, and hence $\psi\mu = 1$.

That is, ψ and μ are reciprocal isomorphisms of R and N , and we have $R \sim N$.

Conversely, if N is a net having the property of the theorem, then, since N is the homomorphic image of a free net F (Theorem 1.4), we have N isomorphic to a subnet of F . But a subnet of a free net is free (Corollary to Theorem 1.5); hence N is a free net.

1.6. Existence of subnets on a countably infinite free set of generators. The following Lemma is needed in the proof of the next theorem:

LEMMA C. *If $J^{(0)} \subset J^{(1)} \subset \dots \subset J^{(i)} \subset \dots$ is a maximal open extension chain generating the net $N = \bigcup_{i \geq 0} J^{(i)}$, and if the sub-half-net H of N is closed in $J^{(2i+1)}$ (for some i), then H generates a subnet K of N , having the following properties:*

$$\text{I. } K \cap J^{(j)} = H \cap J^{(j)}, \quad j \leq 2i + 1.$$

$$\text{II. } (K \cap J^{(2j)})_L = K \cap J^{(2j+1)}, \quad j > i.$$

Proof. Let $H = H_0 \subset H_1 \subset \dots \subset H_i \subset \dots$ be the maximal extension chain of H in N . Then $K = \bigcup_{i \geq 0} H_i$ is the subnet of N which is generated by H (Corollary to Lemma 1.1).

We now prove, by complete induction with respect to j , the validity of the following propositions:

(A) If x is in H_{j+1} , but not in H_j , then x is in $J^{(2i+j+1)}$, but not in $J^{(2i+j)}$, for all $j \geq 0$.

(B) H_{j+1} is closed in $J^{(2i+j+1)}$, for all $j \geq 0$.

(B) is needed only for the inductive proof of (A).

Since H is closed in $J^{(2i+1)}$, every point of H carries one line of each pencil in H , and hence $H = H_1$. Then for $j = 0$, proposition (A) is vacuously true, and, since $H = H_1$ is closed in $J^{(2i+1)}$, proposition (B) is valid.

Suppose that (A) and (B) hold for $j = k$; we distinguish two cases:

Case 1. $k = 2n$; then H_{2n+1} is closed in $J^{(2i+2n+1)}$, and clearly, $H_{2n+2} \subset J^{(2i+2n+2)}$.

If x is in H_{2n+2} , but not in H_{2n+1} , then x is a point on at least two lines of $H_{2n+1} \subset J^{(2i+2n+1)}$. If x were in $J^{(2i+2n+1)}$, then x would be in H_{2n+1} , since H_{2n+1} is closed in $J^{(2i+2n+1)}$. Hence x is in $J^{(2i+2n+2)}$, but not in $J^{(2i+2n+1)}$, and we have established the validity of (A) for $j = k + 1 = 2n + 1$.

Suppose that P is in H_{2n+2} , and P lies on h in $J^{(2i+2n+2)}$. Since $J^{(2i+2n+2)} = (J^{(2i+2n+1)})_P$, h is necessarily in $J^{(2i+2n+1)}$. Then if P is also in $H_{2n+1} \subset J^{(2i+2n+1)}$, h is in H_{2n+1} and hence in H_{2n+2} , by closure condition (a) for H_{2n+1} . If P is not in H_{2n+1} then P is on at least two lines h', h'' , of H_{2n+1} . But, by property (A) just established, then one of these two lines must be h , since $J^{(2i+2n+2)}$ is an open extension. Hence h is in H_{2n+2} .

If k, k' , are in H_{2n+2} , and if k, k' , lie on Q in $J^{(2i+2n+2)}$, then Q is in H_{2n+2} , since in $H_{2n+2} = (H_{2n+1})_P \subset J^{(2i+2n+2)}$, every pair of lines of different pencils meet in a point.

Hence (B) is established for $j = k + 1 = 2n + 1$.

Case 2. $k = 2n + 1$; then H_{2n+2} is closed in $J^{(2i+2n+2)}$, and clearly $H_{2n+3} \subset J^{(2i+2n+3)}$.

If x is in H_{2n+3} , but not in H_{2n+2} , then x is a line on at least one point of $H_{2n+2} \subset J^{(2i+2n+2)}$. If x were in $J^{(2i+2n+2)}$, then x would be in H_{2n+2} , since

H_{2n+2} is closed in $J^{(2i+2n+2)}$. Hence x is in $J^{(2i+2n+3)}$, but not in $J^{(2i+2n+2)}$, and we have established the validity of (A) for $j = k + 1 = 2n + 2$.

If P is in H_{2n+3} , and if P lies on h in $J^{(2i+2n+3)}$, then h is in $H_{2n+3} = (H_{2n+2})_L \subset J^{(2i+2n+3)}$. If h, h' are in H_{2n+3} , and if h, h' meet in a point Q in $J^{(2i+2n+3)}$, then, since $J^{(2i+2n+3)} = (J^{(2i+2n+2)})_L$, Q is necessarily in $J^{(2i+2n+2)}$. Hence, if both h, h' are also in H_{2n+2} , then Q is in H_{2n+2} , by closure condition (b) for H_{2n+2} . Suppose that one of the lines, say h , is not in H_{2n+2} ; then by (A), h is in $J^{(2i+2n+3)}$, but not in $J^{(2i+2n+2)}$, and hence Q must be the one point on h in $H_{2n+3} \subset J^{(2i+2n+3)}$, since $J^{(2i+2n+3)}$ is an open extension. Hence H_{2n+3} is closed in $J^{(2i+2n+3)}$, and (B) is valid for $j = k + 1 = 2n + 2$. This completes the induction.

Now $H \circ J^{(j)}$ is contained in $K \circ J^{(j)}$, certainly. Furthermore, if x is in $K \circ J^{(j)}$, for $j \leq 2i + 1$, then x is in H ; for, if x were *not* in H , then, for some $k \geq 1$, x would be in H_{k+1} , but not in H_k . But then, by proposition (A), x would be in $J^{(2i+k+1)}$, but not in $J^{(2i+k)}$, ($k \geq 1$), which is impossible since x is in $K \circ J^{(2i+1)} \subset J^{(2i+1)}$.

Hence $K \circ J^{(j)}$ is contained in $H \circ J^{(j)}$, for $j \leq 2i + 1$, so that we have finally, $K \circ J^{(j)} = H \circ J^{(j)}$, for $j \leq 2i + 1$, proving I of our lemma.

Furthermore $(K \circ J^{(2j)})_L \subset K \circ (J^{(2j)})_L = K \circ J^{(2j+1)}$, since $K_L = K$. Consider an element x in $K \circ J^{(2j+1)}$, for $j > i$; we have two possibilities:

(i) x is in $J^{(2j)}$; then x is in $K \circ J^{(2j)} \subset (K \circ J^{(2j)})_L$.

(ii) x is not in $J^{(2j)}$; then x is a line on P of $J^{(2j)}$ and on no other point in $J^{(2j+1)}$. x is not in H , since $H \subset J^{(2i+1)}$, and we have $j > i$. Then x is in H_{2k+1} , but not in H_{2k} , for some $k \geq 1$, and x carries a point Q from H_{2k} . But then x is in $J^{(2i+2k+1)}$, not in $J^{(2i+2k)}$, on Q of $J^{(2i+2k)}$, by proposition (A) of this proof. Since x is also in $J^{(2j+1)}$, but not in $J^{(2j)}$, we must have $2j = 2i + 2k$, and $Q = P$. Then $Q = P$ is in $K \circ J^{(2j)}$ and x is in $(K \circ J^{(2j)})_L$.

Hence in both cases we have $K \circ J^{(2j+1)} \subset (K \circ J^{(2j)})_L$, and finally, $(K \circ J^{(2j)})_L = K \circ J^{(2j+1)}$, for $j > i$, proving II of our lemma.

THEOREM 1.7. *If the non-null net N is freely generated by its sub-half-net $J \neq N$, then N contains a free net on a countably infinite free set of generators.*

Proof. N may be represented in the form: $N = \bigcup J^{(i)}$, where $J^{(0)} = J$, and $J^{(i+1)}$ is a complete open (L or P) extension of $J^{(i)}$.

In $J^{(2i+2)}$, for $i > 1$ and sufficiently large, there are at least two lines of each pencil and at least two points on each of these lines, since N is not a

null-net. Since $N \neq J$, we have $J^{(2i-1)} \neq J^{(2i-2)}$, (property (2°) for maximal open extension chains). Then there exists a line x in $J^{(2i-1)}$, but not in $J^{(2i-2)}$, and a point P from $J^{(2i-2)}$ which does not lie on x . It is no loss of generality to assume that x is a t -line, t^* , and that $r(P)$ does not meet t^* in $J^{(2i-1)}$.

We define, then, for the net N , a "ladder," $P-t^*$, as follows: Let $P_0 = t^* \cdot r(P)$; suppose that we have already defined P_{4j} for $j \geq 0$: Then we define:

$$\begin{aligned} s_{4j+1} &= s(P_{4j}); & P_{4j+2} &= s_{4j+1} \cdot t(P); \\ r_{4j+3} &= r(P_{4j+2}); & P_{4j+4} &= r_{4j+3} \cdot t^*. \end{aligned}$$

Note that $P_{4j+2} < t(P)$, and $P_{4j} < t^*$, for all j . See Fig. 1.

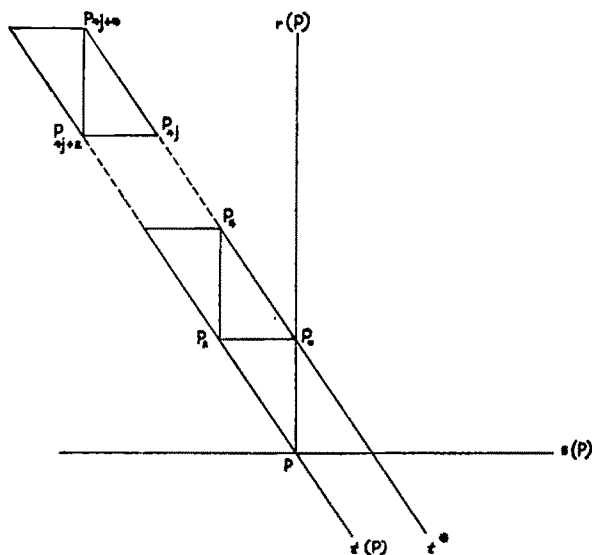


FIGURE 1.

Now we verify the validity of the following statements, for all $j \geq 0$:

- P_{4j} is in $J^{(2i+4j)}$, but not in $J^{(2i+4j-1)}$;
- s_{4j+1} is in $J^{(2i+4j+1)}$, but not in $J^{(2i+4j)}$;
- P_{4j+2} is in $J^{(2i+4j+2)}$, but not in $J^{(2i+4j+1)}$;
- r_{4j+3} is in $J^{(2i+4j+3)}$, but not in $J^{(2i+4j+2)}$.

$P_0 = t^* \cdot r(P)$ is in $J^{(2i)}$, but not in $J^{(2i-1)}$ (since t^* does not meet $r(P)$ in $J^{(2i-1)}$), and P_0 lacks an s -line.

Suppose that P_{4j} is in $J^{(2i+4j)}$, but not in $J^{(2i+4j-1)}$, and that P_{4j} lacks an s -line.

Then $s_{4j+1} = s(P_{4j})$ is in $J^{(2i+4j+1)}$, but not in $J^{(2i+4j)}$, and s_{4j+1} carries P_{4j} and no other point in $J^{(2i+4j+1)}$. Since $P_{4j} < t^*$, and, since P_{4j} , not being in $J^{(2i+4j-1)}$, is distinct from P , s_{4j+1} does not meet $t(P)$ in $J^{(2i+4j+1)}$.

Hence $P_{4j+2} = s_{4j+1} \cdot t(P)$ is in $J^{(2i+4j+2)}$, but not in $J^{(2i+4j+1)}$. Clearly, P_{4j+2} lacks an r -line.

Then $r_{4j+3} = r(P_{4j+2})$ is in $J^{(2i+4j+3)}$, but not in $J^{(2i+4j+2)}$. In $J^{(2i+4j+3)}$, r_{4j+3} carries P_{4j+2} and no other point. Since P_{4j+2} lies on $t(P)$ and hence on no other t -line of N , clearly r_{4j+3} does not meet t^* in $J^{(2i+4j+3)}$.

Then $P_{4j+4} = t^* \cdot r_{4j+3}$ is in $J^{(2i+4j+4)}$, but not in $J^{(2i+4j+3)}$, and P_{4j+4} lacks an s -line.

Hence the validity of the foregoing statements has been proven.

We shall show that the set of lines:

$$K = \{r(P), s(P), s_5, s_{13}, \dots\} = r(P) \cup s(P) \cup \bigcup_{j=0}^{\infty} s_{8j+5}$$

is a free set of generators of a subnet S of N .

First, note that $r(P) \cup s(P) = K(0)$ generates the null-net $S(0)$ on P , and that $S(0)$ is free over $K(0)$. Since $r(P), s(P), t(P)$, are in $J^{(2i-1)}$, we have $S(0) \subset J^{(2i-1)}$. For definiteness, we assume that $S(0) \subset J^{(0)}$ since in later applications of this theorem this will be the case. This assumption does not constitute a restriction on the theorem, but is merely a matter of convenience.

Let $M(1) = S(0) \cup s_5$. Then $M(1)$ is closed in $J^{(2i+5)}$, since $S(0)$ is certainly closed in $J^{(2i+5)}$, and the only lines which meet s_5 in $J^{(2i+5)}$ are the lines t^* and r_3 , neither of which is in $S(0)$. By Lemma C, then $M(1)$ generates a subnet $S(1)$ of N , which has the following properties:

$$\text{I. } S(1) \cap J^{(j)} = M(1) \cap J^{(j)}, \quad j \leq 2i + 5$$

$$\text{II. } (S(1) \cap J^{(2j)})_L = S(1) \cap J^{(2j+1)}, \quad j > i + 2.$$

From Theorem 1.5, then $S(1)$ is freely generated by

$$\begin{aligned} & (S(1) \cap J^{(0)}) \cup \bigcup_{j=0}^{\infty} \{S(1) \cap J^{(2j+1)} - (S(1) \cap J^{(2j)})_L\} \\ &= (S(1) \cap J^{(0)}) \cup \bigcup_{j=0}^{i+2} \{S(1) \cap J^{(2j+1)} - (S(1) \cap J^{(2j)})_L\} \cup \\ & \quad \bigcup_{j > i+2} \{S(1) \cap J^{(2j+1)} - (S(1) \cap J^{(2j)})_L\}, \\ &= (M(1) \cap J^{(0)}) \cup \bigcup_{j=0}^{i+2} \{M(1) \cap J^{(2j+1)} - (M(1) \cap J^{(2j)})_L\}, \end{aligned}$$

by I and II above.

But $M(1) = S(0) \cup s_5$, where $S(0) \subset J^{(0)}$ and s_5 is in $J^{(2i+5)}$, but not in $J^{(2i+4)}$. Hence

$$\begin{aligned} M(1) \cap J^{(j)} &= S(0) \cap J^{(j)} = S(0), \text{ for } j < 2i + 5, \text{ and} \\ M(1) \cap J^{(2i+5)} &= (S(0) \cap J^{(2i+5)}) \cup s_5 = S(0) \cup s_5. \text{ Then} \\ \bigcup_{j=0}^{i+1} \{M(1) \cap J^{(2j+1)} - (M(1) \cap J^{(2j)})_L\} \\ &= \bigcup_{j=0}^{i+1} \{S(0) \cap J^{(2j+1)} - (S(0) \cap J^{(2j)})_L\} \end{aligned}$$

is vacuous, but

$$\{M(1) \cap J^{(2j+5)} - (M(1) \cap J^{(2i+4)})_L\} = (S(0) \cup s_5) - S(0) = s_5.$$

Hence $S(1)$ is freely generated by $S(0) \cup s_5$.

Then $S(1)$ is certainly freely generated by

$$K(1) = r(P) \cup s(P) \cup s_5.$$

Note that, since $S(1) \cap J^{(j)} = M(1) \cap J^{(j)}$, $j \leq 2i + 5$, then t^* is not in $S(1)$, since t^* is in $J^{(2i-1)}$, but not in $M(1) = S(0) \cup s_5$.

Restating the above results in a form more suitable for induction purposes, we have that $S(1)$ is a subnet of N which has a free set of generators: $K(1) = r(P) \cup s(P) \cup s_5$, and which enjoys the following properties:

- (1) $(S(1) \cap J^{(0)}) \cup \bigcup_{j=0}^{i+2} \{S(1) \cap J^{(2j+1)} - (S(1) \cap J^{(2j)})_L\} = S(0) \cup s_5.$
- (2) $(S(1) \cap J^{(2j)})_L = S(1) \cap J^{(2j+1)}, \quad j > i + 2.$
- (3) t^* is not in $S(1).$

Now assume that there exists a subnet $S(k)$ of N , with a free set of generators: $K(k) = r(P) \cup s(P) \cup \bigcup_{j=0}^{k-1} s_{8j+5}$, ($k \geq 1$), enjoying the following properties:

- (1) $(S(k) \cap J^{(0)}) \cup \bigcup_{j=0}^{i+4k-2} \{S(k) \cap J^{(2j+1)} - (S(k) \cap J^{(2j)})_L\} \\ = S(0) \cup \bigcup_{j=0}^{k-1} s_{8j+5}.$
- (2) $(S(k) \cap J^{(2j)})_L = S(k) \cap J^{(2j+1)}, \quad j > i + 4k - 2.$
- (3) t^* is not in $S(k).$

Suppose that s_{8k+5} is in $S(k)$; then, since s_{8k+5} is in $J^{(2i+8k+5)}$, it is in

$S(k) \circ J^{(2i+8k+5)} = (S(k) \circ J^{(2i+8k+4)})_L$. Furthermore, since s_{8k+5} is not in $J^{(2i+8k+4)}$, it is not in $S(k) \circ J^{(2i+8k+4)}$. Hence s_{8k+5} is on a point Q of $S(k) \circ J^{(2i+8k+4)}$. But s_{8k+5} carries P_{8k+4} and no other point of $J^{(2i+8k+4)}$. Hence $Q = P_{8k+4}$ is in $S(k)$; but P_{8k+4} lies on l^* , and hence l^* is in $S(k)$, which is impossible. Hence s_{8k+5} is not in $S(k)$.

Moreover, r_{8k+3} is not in $S(k)$, by the same line of reasoning as that used above.

Clearly $S(k) \circ J^{(2i+8k+5)}$ is closed in $J^{(2i+8k+5)}$, since $S(k)$ is a net. Then $M(k+1) = (S(k) \circ J^{(2i+8k+5)}) \cup s_{8k+5}$ is closed in $J^{(2i+8k+5)}$, since the only lines which meet s_{8k+5} in $J^{(2i+8k+5)}$ are l^* and r_{8k+3} , neither of which is in $S(k)$.

Then $M(k+1)$ generates a net $S(k+1) \subset N$, with properties:

- I. $S(k+1) \circ J^{(j)} = M(k+1) \circ J^{(j)}$, $j \leq 2i+8k+5$.
- II. $(S(k+1) \circ J^{(2j)})_L = S(k+1) \circ J^{(2j+1)}$, $j > i+4k+2$,

by Lemma C.

Then, by Theorem 1.5, $S(k+1)$ is freely generated by:

$$\begin{aligned} & (S(k+1) \circ J^{(0)}) \circ \bigcup_{j=0}^{\infty} \{S(k+1) \circ J^{(2j+1)} - (S(k+1) \circ J^{(2j)})_L\} \\ &= (S(k+1) \circ J^{(0)}) \cup \bigcup_{j=0}^{i+4k+2} \{S(k+1) \circ J^{(2j+1)} - (S(k+1) \circ J^{(2j)})_L\} \\ & \quad \cup \bigcup_{j > i+4k+2} \{S(k+1) \circ J^{(2j+1)} - (S(k+1) \circ J^{(2j)})_L\} \\ &= (M(k+1) \circ J^{(0)}) \bigcup_{j=0}^{i+4k+2} \{M(k+1) \circ J^{(2j+1)} - (M(k+1) \circ J^{(2j)})_L\}, \end{aligned}$$

by I and II above

But $M(k+1) = (S(k) \circ J^{(2i+8k+5)}) \cup s_{8k+5}$, where s_{8k+5} is in $J^{(2i+8k+5)}$, but not in $J^{(2i+8k+4)}$, so that

$$M(k+1) \circ J^{(j)} = S(k) \circ J^{(j)} \text{ for } j < 2i+8k+5, \text{ and}$$

$$M(k+1) \circ J^{(2i+8k+5)} = (S(k) \circ J^{(2i+8k+5)}) \cup s_{8k+5}.$$

Since $(S(k) \circ J^{(2j)})_L = S(k) \circ J^{(2j+1)}$, for $j > i+4k-2$, we have that

$$\begin{aligned} & \bigcup_{j=i+4k-1}^{i+4k+1} \{M(k+1) \circ J^{(2j+1)} - (M(k+1) \circ J^{(2j)})_L\} \\ &= \bigcup_{j=i+4k-1}^{i+4k+1} \{S(k) \circ J^{(2j+1)} - (S(k) \circ J^{(2j)})_L\} \end{aligned}$$

is vacuous.

However,

$$\begin{aligned}
& \{M(k+1) \cap J^{(2i+8k+5)} - (M(k+1) \cap J^{(2i+8k+4)})_L\} \\
&= \{(S(k) \cap J^{(2i+8k+5)}) \cup s_{8k+5} - (S(k) \cap J^{(2i+8k+4)})_L\} \\
&= \{(S(k) \cap J^{(2i+8k+5)}) \cup s_{8k+5} - (S(k) \cap J^{(2i+8k+5)})\} = s_{8k+5}.
\end{aligned}$$

Hence $S(k+1)$ is freely generated by

$$\begin{aligned}
& (S(k) \cap J^{(0)}) \cup \bigcup_{j=0}^{i+4k-2} \{S(k) \cap J^{(2j+1)} - (S(k) \cap J^{(2j)})_L\} \cup s_{8k+5} \\
&= S(0) \cup \bigcup_{j=0}^{k-1} s_{8j+5} \cup s_{8k+5} = S(0) \cup \bigcup_{j=0}^k s_{8j+5}.
\end{aligned}$$

Then $S(k+1)$ is certainly freely generated by

$$K(k+1) = r(P) \cup s(P) \cup \bigcup_{j=0}^k s_{8j+5}.$$

Hence, we have an ascending chain of subnets:

$$S(0) \subset S(1) \subset \cdots \subset S(k) \subset S(k+1) \subset \cdots \subset N,$$

such that $S(i)$ is freely generated by $K(i) = r(P) \cup s(P) \cup \bigcup_{j=0}^{i-1} s_{8j+5}$, and $K(i) \subset K(i+1)$, for every i . Then it is easily verified that $S = \bigcup_{i=0}^{\infty} S(i)$ is a subnet of N which is freely generated by

$$K = \bigcup_{i=0}^{\infty} K(i) = r(P) \cup s(P) \cup \bigcup_{j=0}^{\infty} s_{8j+5}.$$

That is, S is a free subnet of N having a free set of generators, K , countably infinite in number.

1.7. Centered half-nets. We now consider a special class of half-nets which will find their direct counterpart in loop theory in the next section.

Definition. The half-net T is termed P -centered, for P a point in T , if the following conditions are satisfied:

- (a) If Q is in T , then Q carries $r(Q)$, $s(Q)$, $t(Q)$, in T .
- (b) $s(P)$ meets every x -line in T , for $x \neq s$.
- $r(P)$ meets every x -line in T , for $x \neq r$.

Note that a net is P -centered for P , any point in the net.

THEOREM 1.8. Given a net R and a sub-half-net M of R , containing the

point P , then there exists a uniquely determined P -centered half-net, $M^* = (M, P)$, such that M^* is generated by M , and no proper sub-half-net of M^* is a P -centered half-net containing M .

Proof. There exist P -centered sub-half-nets of the net R ; for instance, R itself. Consider the cross-cut, M^* , of all P -centered sub-half-nets of R which contain M . Now it is easily verified that M^* is a P -centered sub-half-net of R which contains M , and that there does not exist any proper P -centered sub-half-net of M^* containing M . Suppose that S is a sub-half-net such that $M \subset S \subset M^*$, with S closed in M^* . Then clearly, S , too, is P -centered, and hence $S = M^*$, so that M^* is generated by M and thus meets all requirements.

We shall use throughout the notation, (M, P) , to represent the smallest P -centered half-net containing the half-net M , and generated by M .

COROLLARY. *If the net R is free over and generated by its sub-half-net J , and if J contains the point P , then R is freely generated by (J, P) .*

This is an immediate consequence of Theorem 1.8, and of Lemma 1.3,(1).

Definition. If the net N is freely generated by its P -centered sub-half-net, (J, P) , and if (J, P) has the following property:

(P_f) If Q is in (J, P) , then Q is on $r(P)$ or on $s(P)$,

then we define (J, P) as a P -centered free set of generators of N . A P -centered free net, then, is a net having a P -centered free set of generators.

THEOREM 1.9. *A net N has a P -centered free set of generators (for some point P in N) if, and only if, N has a free set of generators.*

Proof. If N has a P -centered free set of generators (J, P) , then it is obvious that the sub-half-net K , consisting of all the r -lines in (J, P) , together with the s -line, $s(P)$, constitutes a free set of generators of N .

Suppose that N has a free set of generators, M . Since M consists entirely of lines, in the maximal open extension chain of M , we have $M = M^{(1)}$, but $M^{(1)} \neq M^{(2)}$. Let P be any one of the points adjoined in $M^{(2)}$; then P lies on two and only two lines, say h, h' , from M . Let M^* be the sub-half-net of N which is formed from M by deleting the lines h, h' , and substituting, instead, the point $P = h \cdot h'$ in N . It is clear that N is freely generated by M^* , since $M^{(2)}$ is free over and generated by M^* . Then N is freely generated by (M^*, P) , the P -centered sub-half-net of N which is determined by M^* , (Corollary to Theorem 1.8). Furthermore, (M^*, P) is a P -centered free set of generators of N , since if there were a point Q in

(M^*, P) , such that Q did not lie on $r(P)$ nor on $s(P)$, then Q would have three lines, one of each pencil, in (M^*, P) . At most two of these lines would be lines of M , since M contains no points, and N is free over M . Then it is almost obvious, from the properties of P -centered half-nets, that there would be a P -centered half-net containing M and having but two of these lines, hence being properly contained in (M^*, P) , which is impossible.

Definition. A net N containing the point P , is the P -centered free sum of its P -centered subnets, $(M(v), P)$, if the following conditions are satisfied:

(i) N is generated by $\bigcup_v (M(v), P)$.

(ii) If $\phi(v)$ are homomorphisms of $(M(v), P)$ into a net R , with $P\phi(v) = P\phi(w)$, for all v, w , then there exists a homomorphism ϕ of N , which coincides with $\phi(v)$ on $(M(v), P)$, for each v .

One verifies that if N is the P -centered free sum of its subnets, $(M(v), P)$, then $(M(v), P) \circ (M(w), P)$ consists of the null-net on P , for all $v \neq w$.

Note. If N is the P -centered free sum of its subnets, $(M(v), P)$, then N is the P -centered free sum of $(M(v), P)$ and the net generated by $\bigcup_{w \neq v} (M(w), P)$.

THEOREM 1.10. *If the net N containing the point P , is the free sum of its subnets, (K, P) and S , where (K, P) is generated by (K_0, P) , and S is generated by S_0 , then N is the P -centered free sum of the nets (K, P) and (T, P) , where (T, P) is generated by (T_0, P) , and where T_0 is the half-net which is obtained from S_0 by adjoining the point P .*

Proof. Let σ and τ be homomorphisms of (K, P) and (T, P) , respectively, such that $P\sigma = P\tau$; clearly, τ induces a homomorphism ρ in S . There exists one and only one homomorphism ϕ of N which induces σ in (K, P) and ρ in S , since $N = (K, P) * S$. Thus ϕ coincides with the homomorphism τ of (T, P) in the point P and in S . But (T, P) is generated by $T_0 = S \vee P$ (since (T_0, P) is generated by T_0 , and (T, P) is generated by (T_0, P)), implying that ϕ coincides with τ everywhere in (T, P) , (Theorem 1.1), as was to be shown.

2. HALF-LOOPS AND CENTERED HALF-NETS.

2.1. Representation of half-loops by centered half-nets. A half-loop H is a set of elements with one composition, called addition, $a + b$, subject to the following rules:

- (i) If $a + b = c$, and $a + b = d$, then $c = d$.
- (ii) If $a + d = a + f$, then $d = f$; if $b + a = c + a$, then $b = c$.
- (iii) There exists a null element, 0 , satisfying

$$a + 0 = 0 + a = a, \text{ for every } a \text{ in } H.$$

The term, "half-loop," is not to be confused with "semi-loop," which is defined as follows: A *semi-loop*, S , is a half-loop which satisfies the additional condition:

(i'). If a, b , are in S , then there is at least one element c in S , such that $a + b = c$.

A *loop*, L , is a semi-loop which satisfies the further condition:

(ii'). If a, b , are in L , then there is at least one x , and at least one y in L , such that $a + x = b$, and $y + a = b$.

This definition is equivalent to the usual definition of a loop.⁵

With every half-loop T we associate an O -centered half-net, (M, O) , as follows:

Let the points of (M, O) be ordered pairs of elements, (a, b) , from T , subject to the following condition:

(P). (a, b) is a point in (M, O) , if and only if, there is at least one element c in T , such that $a + b = c$.

Thus (o, o) is in (M, O) , and we let $O = (o, o)$.

With every element b in T , we associate one and only one r -line, $r(b)$; one and only one s -line, $s(b)$; and one and only one t -line, $t(b)$. The point (x, y) in (M, O) is on $r(b)$, if and only if $x = b$; it is on $s(b)$, if and only if, $y = b$; and (x, y) of (M, O) is on $t(b)$, if and only if $x + y = b$.

In particular, if T is the null loop—i. e., if T consists of the null element alone, then (M, O) is the null net on O .

LEMMA 2.1. (M, O) , as defined above, is an O -centered half-net.

Proof. We verify immediately, as direct consequences of the properties of the half-loop T and condition (P), that (M, O) is a half-net in which every point has one and only one line of each pencil. Also, $r(a)$ meets $s(O)$ in (a, o) ; $s(a)$ meets $r(O)$ in (o, a) ; and $t(a)$ meets $r(O)$ and $s(O)$ in (o, a) and (a, o) , respectively, so that (M, O) satisfies the further conditions for an O -centered half-net.

⁵ See Baer (4).

We denote the dependence of (M, O) upon the half-loop T , by writing $(M, O) = N(T)$, where the letter " N " is used in the generic sense to signify half-net or net. In using this notation, we understand that the associated half-net is centered on the point $O = (o, o)$, and we reserve the letter " O " from now on for this role.

Now one verifies similarly, the

COROLLARY. *If T is a loop, then $N(T)$ is a net.*

This is a well-known theorem, of which our lemma is a generalization.⁶

LEMMA 2.2. *Given an O -centered half-net, (M, O) , then there is a representation of (M, O) as $N(T)$, where T is a half-loop.*

Proof. To each point on the line, $s(O)$, we assign a symbol, letting " o " be the symbol for the point O . We wish to show that this set of symbols, T , forms a half-loop under addition as defined below:

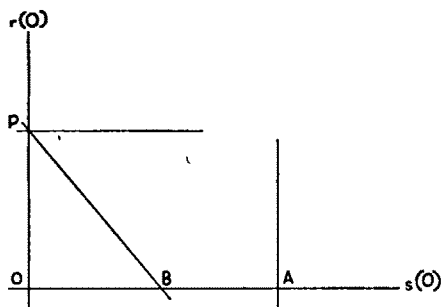


FIGURE 2.

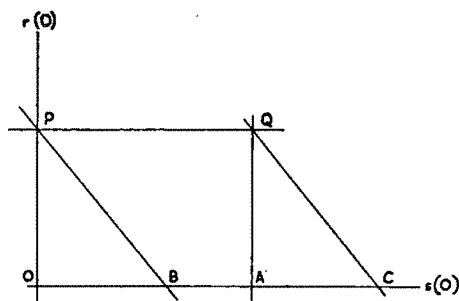


FIGURE 3.

Let A and B be points on $s(O)$ with symbols a and b , respectively. Then $t(B)$ and $r(A)$ are in (M, O) , and $t(B)$ meets $r(O)$ in a point P . Furthermore, $s(P)$ is in (M, O) . We proceed as follows:

(i) If $r(A)$ and $s(P)$ do not meet in (M, O) , then we say there is no symbol in T representing $a + b$, or, simply that $a + b$ is not defined in T . (See Fig. 2.)

(ii) If $r(A)$ and $s(P)$ meet in Q , then $t(Q)$ is in (M, O) , and $t(Q)$ meets $s(O)$ in a point C . Let the symbol for C be c . Then we define c to be $a + b$. (See Fig. 3.)

That the addition of symbols in T , as defined above, is single-valued,

⁶ See Bol (1).

follows from the fact that each of the points and lines used to define $a + b$ is well-determined by the properties of (M, O) . Hence, for elements of T , $a + b = c$, and $a + b = d$, imply that $c = d$. Clearly, the symbol " o " is the null element in this addition, and the two cancellation laws for T are easily verified. Hence T is a half-loop.

Returning to Fig. 3, if we assign to the point Q , the "coordinates" (a, b) , we can show that the points of M may be mapped in a 1 — 1 way upon the ordered pairs of symbols in T . Then points on the same r -line will have the same first coordinate, points on the same s -line will have their second coordinates the same, and points on the same t -line will be such that the sum of their two coordinates is the same for all points on the line. A point P will be in (M, O) and will have coordinates, (a, b) , if and only if, $a + b$ is defined by the net addition described above. Hence we have (M, O) represented as $N(T)$, where T is a half-loop whose structure is completely determined by (M, O) . We denote the dependence of the half-loop T upon the O -centered half-net, (M, O) , by writing $T = L(M, O)$, where the letter " L " is used in the generic sense to indicate loop or half-loop.

One verifies, similarly, the following:

COROLLARY. *If (M, O) is a net, then $L(M, O)$ is a loop.*

This is a well-known theorem, of which our lemma is a generalization.⁷

The following relationships between half-loops and centered half-nets are easily verified, and are here given without proof, numbered for future reference.

1°. If K is a proper sub-half-loop of the half-loop L , then $N(K)$ is a proper sub-half-net of the half-net $N(L)$.

2°. If (M, O) is a proper sub-half-net of the half-net, (R, O) , then $L(M, O)$ is a proper sub-half-loop of the half-loop, $L(R, O)$.

3°. If K and T are sub-half-loops of the half-loop L , then $N(K \circ T) = N(K) \circ N(T) \subset N(L)$.

2.2. Homomorphisms of half-loops and their associated half-nets.

Definition. A homomorphism η of a half-loop K , (into a half-loop, T), is a single-valued mapping of K which preserves sums—that is, if $a + b$ is defined in K , then so is $a\eta + b\eta$, and $(a + b)\eta = a\eta + b\eta$.

LEMMA 2.3. (a) *Given a homomorphism η of the half-loop K into the*

⁷ See Bol (1).

half-loop T , then mapping (x, y) in $N(K) = (M, O)$ upon $(x, y)\phi = (x\eta, y\eta)$ in $N(T) = (R, O')$, constitutes a homomorphism of $N(K)$ into $N(T)$ with $O' = O\phi$.

(b) Given a homomorphism ϕ of (M, O) into (R, O') , such that $O' = O\phi$, then ϕ is induced by a homomorphism η of $K = L(M, O)$ into $T = L(R, O')$ in the manner indicated in (a).

Proof. (a) If (x, y) is in $N(K)$, then there is a z in K , such that $x + y = z$. Then $x\eta, y\eta$, are in T , with $x\eta + y\eta = (x + y)\eta = z\eta$ in T . Hence $(x\eta, y\eta)$ is a point in $N(T)$, and the mapping ϕ is a single-valued mapping of points of $N(K)$ upon points of $N(T)$. We have $O = (o, o)$ mapped under ϕ on $(o\eta, o\eta) = (o', o') = O'$, where o' is the null element in T .

Now (a, y) and (a, z) in $N(K)$ are mapped by ϕ upon $(a\eta, y\eta)$ and $(a\eta, z\eta)$, respectively. Hence points on the same r -line in $N(K)$ are mapped on points on the same r -line in $N(T)$. Similarly, (x, b) and (y, b) on the same s -line in $N(K)$ are mapped upon $(x\eta, b\eta)$ and $(y\eta, b\eta)$ on the same s -line in $N(T)$. Consider points (x, y) and $(x + y, o)$ on the same t -line in $N(K)$. $(x, y)\phi = (x\eta, y\eta)$, and $(x + y, o)\phi = [(x + y)\eta, o']$. But $(x + y)\eta = x\eta + y\eta$, and hence, $(x + y, o)\phi = (x\eta + y\eta, o')$. That is, points on the same t -line in $N(K)$ are mapped on points on the same t -line in $N(T)$.

Then ϕ is a homomorphism of $N(K)$ into $N(T)$, and we have proved (a).

(b) (M, O) determines $K = L(M, O)$, and hence we may write $(M, O) = N(K)$; similarly, (R, O') determines $T = L(R, O')$, and we have $(R, O') = N(T)$.

We define a single-valued mapping η as follows: $(x, o)\phi = (x\eta, o')$. Now (o, x) and (o, o) are on the same r -line in (M, O) ; then $(o, x)\phi$ and $(o, o)\phi = (o', o')$ must be on the same r -line in (R, O') . Hence $(o, x)\phi = (o', y)$, for some y in T .

(x, o) and (o, x) are on the same t -line in (M, O) ; then $(x, o)\phi = (x\eta, o')$ and $(o', x)\phi = (o', y)$ are on the same t -line in (R, O') . Then $x\eta + o' = o' + y$, or $y = x\eta$. Hence $(o, x)\phi = (o', x\eta)$.

Finally, we have $(x, y)\phi = (x\eta, y\eta)$, since $(x, y)\phi = (w, z)$ and $(x, o)\phi = (x\eta, o')$ on the same r -line in (R, O') implies that $w = x\eta$, and $(x, y)\phi$ and $(o, y)\phi = (o', y\eta)$ on the same s -line, implies that $z = y\eta$.

Now consider (x, y) and $(x + y, o)$ on the same t -line in (M, O) ; then $(x, y)\phi = (x\eta, y\eta)$, and $(x + y, o)\phi = [(x + y)\eta, o']$ must be on the same t -line in (R, O') . Hence $(x + y)\eta = x\eta + y\eta$; that is, η is a homomorphism of $K = L(M, O)$ into $T = L(R, O')$.

2.3. Concepts of freeness for half-loops and their associated half-nets.

A half-loop L is *free over* its sub-half-loop K if every homomorphism of K may be extended to a homomorphism of L .

Definition. The sub-half-loop H is *closed* in the half-loop K , if $H \subset K$, and if, given any two of the elements a, b, c , in H with $a + b = c$ in K , then the third element is in H .

The half-loop K is *generated* by its sub-half-loop S , if there is no proper sub-half-loop H of K which contains S and is closed in K .

It is easily verified that the half-loop K is generated by its sub-half-loop S , if and only if, $N(K)$ is generated by $N(S)$.

If the half-loop K is free over and generated by its sub-half-loop S , we shall say that K is *freely generated* by S .

LEMMA 2.4. *The half-loop T is freely generated by its sub-half-loop K , if and only if, $N(T)$ is freely generated by $N(K)$.*

Proof. Given a homomorphism ϕ of $N(K)$ into a net R , then R is $O\phi$ -centered; let $O' = O\phi$, and let $H = L(R, O')$. Now H is a loop and there is a homomorphism η of K into H , which induces ϕ by Lemma 2.3 (b). Since the half-loop T is free over K , η may be extended to a homomorphism ρ of T into H . By Lemma 2.3, (a), then ρ induces a homomorphism ψ of $N(T)$ into $N(H) = R$, and ψ induces ϕ in $N(K)$. Hence $N(T)$ is free over $N(K)$. Since T is generated by K , then $N(T)$ is generated by $N(K)$, proving the Lemma in one direction.

Conversely, if $N(T)$ is freely generated by $N(K)$, by similar arguments we show that any homomorphism of K into a loop S may be extended to a homomorphism of T into S , so that T is free over K ; and T is generated by K , since $N(T)$ is generated by $N(K)$.

Although, for the most part, we shall not be interested in an explicit representation for the elements of a loop in terms of its generators, some such representation is needed in the next section for our formulation of the Subloop Theorem. Accordingly, we define the following chain of half-loops:

Definition. If the loop R is generated by its sub-half-loop T , then R contains the following *generating chain* of half-loops:

$$T \subset T_1 \subset T_2 \subset T_3 \subset \cdots \subset T_i \subset \cdots \subset R,$$

where T_{i+1} is the set of all elements z, x, y , in R , such that: $z = a + b$, $x + a = b$, $a + y = b$, for a, b , in T_i . Furthermore, it is easily verified that

$$R = \bigcup_{i=0}^{\infty} T_i.$$

Note. A direct development of the loop theory of Section 3 may be effected by a study of these generating chains and properly defined free generating chains for half-loops. In such a development, it is true, however, that some of the loop constructions analogous to our net constructions seem relatively more complicated.

The following lemma serves to make more specific the relationship between the elements in the generating chain of a loop R which is freely generated by its half-loop T , and our free net construction:

LEMMA 2.5. *If R is a loop which is freely generated by its half-loop T , and if $T \subset T_1 \subset T_2 \subset \cdots \subset T_i \subset \cdots \subset R$ is the generating chain of T in R , then the following conditions are satisfied, for every $i \geq 0$:*

(L_z). For every z in T_{i+1} , not in T_i , such that $z = a + b$ for a, b , elements in T_i , there exists an integer $j \geq 0$, such that $r(a), s(b)$, are in $N(T)^{(2j+1)}$, but the point (a, b) is not in $N(T)^{(2j)}$.

(L_x). For every x in T_{i+1} , but not in T_i , such that $x + a = b$, for a, b , in T_i , there exists an integer $k \geq 0$ such that $s(a), t(b)$, are in $N(T)^{(2k+1)}$, but the point (x, a) is not in $N(T)^{(2k+1)}$.

(L_y). For every y in T_{i+1} , but not in T_i , such that $a + y = b$, for a, b , in T_i , there exists an integer $l \geq 0$, such that $r(a), t(b)$, are in $N(T)^{(2l+1)}$, but the point (a, y) is not in $N(T)^{(2l+1)}$.

Proof. Since R is freely generated by T , then $N(R) = \bigcup_{j=0}^{\infty} N(T)^{(j)}$ is freely generated by $N(T)$, and hence $N(T)^{(j+1)}$ is an open L or P extension of $N(T)^{(j)}$.

The conditions (L_z), (L_x), (L_y), are clearly satisfied for $i = 0$.

Assume that these conditions hold for all $i \leq M$. We prove, first, before proceeding with the induction, the validity of the following statement:

(K). Under the induction hypothesis, if c is any element in T_{i+1} , but not in T_i , ($i \leq M$), then the following possibilities are exhaustive for the lines $r(c)$, $s(c)$, and $t(c)$ in $N(R)$:

I. There exists an integer $m \geq 1$, such that $r(c)$ is in $N(T)^{(2m+1)}$, but not in $N(T)^{(2m)}$, and such that there exists one and only one point in $N(T)^{(2m+1)}$ on $r(c)$, namely (c, u) , with $u \neq c$, and $u, c + u$, in T_{i+1} .

II. There exists an integer $n \geq 1$, such that $s(c)$ is in $N(T)^{(2n+1)}$, but not in $N(T)^{(2n)}$, and such that there exists one and only one point in $N(T)^{(2n+1)}$ on $s(c)$, namely (v, c) , with $v \neq c$, and $v, v + c$ in T_{i+1} .

III. There exists an integer $p \geq 1$, such that $t(c)$ is in $N(T)^{(2p+1)}$, but not in $N(T)^{(2p)}$, and such that there exists one and only one point in $N(T)^{(2p+1)}$ on $t(c)$, namely (w, w') , with $w, w' \neq c$, and $w, w', w + w'$ in T_{i+1} .

Proof of statement (K). Since c is in T_{i+1} , but not in T_i , there are three possibilities for c : (1) $c = w + w'$; w, w' , in T_i ; (2) $c + u = u'$; u, u' , in T_i ; (3) $v + c = v'$; v, v' , in T_i . It is necessary to consider each of these possibilities separately:

Case (1). $c = w + w'$; since c is not in T_i , then $w, w' \neq c$. By the induction hypothesis, there exists an integer j , such that

$$r(w), s(w'), \in N(T)^{(2j+1)}, \text{ but } (w, w') \notin N(T)^{(2j)}.$$

Then: $(w, w') = r(w) \cdot s(w')$ is in $N(T)^{(2j+2)}$, but not in $N(T)^{(2j+1)}$, and $r(w), s(w')$, are the only lines on (w, w') in $N(T)^{(2j+2)}$.

$t(c) = t(w, w')$ is in $N(T)^{(2j+3)}$, but not in $N(T)^{(2j+2)}$, and (w, w') is the only point in $N(T)^{(2j+3)}$ on $t(c)$, exhibiting the possibility III of statement (K).

Since $(w, w') \neq (c, o)$ and $(w, w') \neq (o, c)$, we have

$$\begin{aligned} (c, o) &= t(c) \cdot s(o) \\ (o, c) &= t(c) \cdot r(o) \end{aligned} \quad \text{in } N(T)^{(2j+4)}, \text{ but not in } N(T)^{(2j+3)}.$$

Then $r(c) = r(c, o)$ is in $N(T)^{(2j+5)}$, but not in $N(T)^{(2j+4)}$, and (c, o) is the only point on $r(c)$ in $N(T)^{(2j+5)}$ —that is, we have the situation in I of statement (K).

Also, $s(c) = s(o, c)$ is in $N(T)^{(2j+5)}$, but not in $N(T)^{(2j+4)}$, and (o, c) is the only point on $s(c)$ in $N(T)^{(2j+5)}$, (see II, under (K)).

Case (2). $c + u = u'$; since c is not in T_i , then $u, u' \neq c$.

By the induction hypothesis, there exists an integer k , such that $s(u), t(u')$, are in $N(T)^{(2k+1)}$, but the point (c, u) is not in $N(T)^{(2k+1)}$.

Then $(c, u) = s(u) \cdot t(u')$ is in $N(T)^{(2k+2)}$, but not in $N(T)^{(2k+1)}$, and $s(u), t(u')$, are the only lines on (c, u) in $N(T)^{(2k+2)}$.

$r(c) = r(c, u)$ is in $N(T)^{(2k+3)}$, but not in $N(T)^{(2k+2)}$, and (c, u) is the only point in $N(T)^{(2k+3)}$ on $r(c)$. (See I under (K)). $(c, u) \neq (c, o)$, since $u' \neq c$ implies that $u \neq o$.

Then $(c, o) = r(c) \cdot s(o)$ is in $N(T)^{(2k+4)}$, but not in $N(T)^{(2k+3)}$, and $r(c), s(o)$, are the only lines on (c, o) in $N(T)^{(2k+4)}$.

$t(c)$ is in $N(T)^{(2k+5)}$, but not in $N(T)^{(2k+4)}$, and (c, o) is the only point in $N(T)^{(2k+5)}$ on $t(c)$. (See III under (K)). Since $c \neq o$, then $(o, c) \neq (c, o)$, and we have: $(o, c) = t(c) \cdot r(o)$ in $N(T)^{(2k+6)}$, but not in $N(T)^{(2k+5)}$, such that $t(c), r(o)$, are the only lines in $N(T)^{(2k+6)}$ on (o, c) .

$s(c) = s(o, c)$ is in $N(T)^{(2k+7)}$, but not in $N(T)^{(2k+6)}$, and (o, c) is the only point on $s(c)$ in $N(T)^{(2k+7)}$. (See II under (K)).

Case (3). $v + c = v'$; since c is not in T_i , then $v, v' \neq c$.

By the induction hypothesis, there exists an integer l , such that $r(v), l(v')$, are in $N(T)^{(2l+1)}$, but the point (v, c) is not in $N(T)^{(2l+1)}$.

Then $(v, c) = r(v) \cdot t(v')$ is in $N(T)^{(2l+2)}$, but not in $N(T)^{(2l+1)}$, and $r(v), t(v')$, are the only lines in $N(T)^{(2l+2)}$ on (v, c) .

$s(c) = s(v, c)$ is in $N(T)^{(2l+3)}$, but not in $N(T)^{(2l+2)}$, and (v, c) is the only point in $N(T)^{(2l+3)}$ on $s(c)$. (See II under (K)).

$(v, c) \neq (o, c)$, since $v' \neq c$ implies that $v \neq o$.

Then $(o, c) = s(c) \cdot r(o)$ is in $N(T)^{(2l+4)}$, but not in $N(T)^{(2l+3)}$, and $s(c), r(o)$, are the only lines in $N(T)^{(2l+4)}$ on (o, c) .

$t(c) = t(o, c)$ is in $N(T)^{(2l+5)}$, but not in $N(T)^{(2l+4)}$, and (o, c) is the only point of $N(T)^{(2l+5)}$ which is on $t(c)$. (See III, (K)).

Since $c \neq o$, then $(o, c) \neq (c, o)$, and we have

$(c, o) = t(c) \cdot s(o)$ in $N(T)^{(2l+6)}$, but not in $N(T)^{(2l+5)}$, such that $t(c), s(o)$, are the only lines in $N(T)^{(2l+6)}$ on (c, o) .

$r(c) = r(c, o)$ is in $N(T)^{(2l+7)}$, but not in $N(T)^{(2l+6)}$, and (c, o) is the only point in $N(T)^{(2l+7)}$ on $r(c)$. (See I under (K)).

This completes the verification of statement (K) , and we are ready to proceed with the inductive proof of the Lemma.

By the induction hypothesis, conditions $(L_z), (L_x), (L_y)$, hold for all $i \leq M$. We now show that they hold for $i = M + 1$, which will complete the proof of the Lemma. We consider, first, condition (L_z) : Given z , an element in T_{M+2} , but not in T_{M+1} , which is of the form $z = a + b$, where a, b , are elements of T_{M+1} , then we wish to show the existence of an integer j , such that $r(a), s(b)$ are in $N(T)^{(2j+1)}$, but the point (a, b) is not in $N(T)^{(2j+1)}$. Since z is not in T_{M+1} , then $a, b, \neq z$, and $a, b, \neq o$. Not both a, b , are in T , since then $z = a + b$ is in $T_1 \leq T_{M+1}$.

First, suppose that both a and b are not in T ; then

(1) a is in T_{i+1} , but not in T_i , for some $i \leq M$, and by statement (K), there exists an integer m , such that $r(a)$ is in $N(T)^{(2m+1)}$, but not in $N(T)^{(2m)}$, and such that there exists one and only one point in $N(T)$ on $r(a)$, namely (a, u) , with $u \neq a$, and $u, a + u$ in T_{i+1} . Now $(a, u) \neq (a, b)$, since if $(a, u) = (a, b)$, then $z = a + b = a + u$ is in $T_{i+1} \subset T_{M+1}$, which is impossible. Hence $r(a)$ is in $N(T)^{(2m+1)}$, but not in $N(T)^{(2m)}$, and $P \neq (a, b)$ is the only point on $r(a)$ in $N(T)^{(2m+1)}$; that is, (a, b) is not in $N(T)^{(2m+1)}$.

(2) b is in T_{j+1} , but not in T_j , for some $j \leq M$, and by statement (K), there exists an integer n , such that $s(b)$ is in $N(T)^{(2n+1)}$, but not in $N(T)^{(2n)}$, and such that there exists one and only one point in $N(T)^{(2n+1)}$ on $s(b)$, namely (v, b) , with $v \neq b$, and $v, v + b$ in T_{j+1} . Now $(v, b) \neq (a, b)$, since if $(v, b) = (a, b)$, then $z = a + b = v + b$ would be in $T_{j+1} \subset T_{M+1}$. Hence $s(b)$ is in $N(T)^{(2n+1)}$, but not in $N(T)^{(2n)}$, and $Q \neq (a, b)$ is the only point on $s(b)$ in $N(T)^{(2n+1)}$; that is, (a, b) is not in $N(T)^{(2n+1)}$.

Let $J = \max(m, n)$; then $r(a)$ and $s(b)$ are in $N(T)^{(2J+1)}$, but the point (a, b) is not in $N(T)^{(2J+1)}$, and our induction carries through in the case both a and b are not in T .

Suppose that a is in T , but b is not in T . Then, clearly, $r(a)$ is in $N(T) \subset N(T)^{(2n+1)}$, so that both $r(a), s(b)$ are in $N(T)^{(2n+1)}$, but (a, b) is not in $N(T)^{(2n+1)}$.

Finally, if b is in T , but a is not in T , then $s(b)$ is in $N(T) \subset N(T)^{(2n+1)}$ so that both $r(a)$ and $s(b)$ are in $N(T)^{(2m+1)}$, but (a, b) is not in $N(T)^{(2m+1)}$.

Thus we have carried through the induction in showing that condition (L_z) also holds for $i = M + 1$. The induction in proving the conditions (L_x) and (L_y) , for $i = M + 1$ goes through by similar arguments, and this concludes the proof of Lemma 2.5.

Definition. The loop L is the *free sum* of its subloops $L(v)$, (we write $L = \sum^* L(v)$), if the following conditions are satisfied:

(1) L is generated by $\bigcup L(v)$.

(2) If $\eta(v)$ are homomorphisms of the $L(v)$ into a loop K , then there is a homomorphism η of L into K which coincides with $\eta(v)$ in $L(v)$.

The following statement of the relationship between free sums of loops and free centered sums of nets is easily verified by means of preceding lemmas:

4°. The loop T is the free sum of its subloops, $T(v)$, if and only if, $N(T)$ is the O -centered free sum of its (O -centered) subnets, $N(T(v))$.

Definition. If the loop T is freely generated by its sub-half-loop K , and if K has the following property:

(f) If a, b , are in K , then $a + b$ exists in K , if and only if, at least one of the elements a, b , is o , then K is defined as a *free set of generators* of T . A *free loop*, then, is a loop which has a free set of generators.

Remark. This definition differs from the usual one (in group theory), in so far as we include o in the free set of generators, whereas o is usually excluded. This has technical advantages, but does not make any real difference.

Note that if $T = \sum^* T(v)$, and if the $T(v)$ are free loops, then T is free. The following statement is easily demonstrated:

5°. If the loop T has a free set of generators, K , then $N(T)$ has an O -centered free set of generators, $N(K)$, and conversely.

3. LOOPS.

We are now able to "translate" directly into loop theory the findings of preceding sections. At the same time, it will be of interest to note comparisons with classical results from the theory of free groups and free Abelian groups. That our definitions of free loop, free sums of loops, etc., may be formulated to embrace the usual definitions of free group, free Abelian group, free products of groups, and direct sums of Abelian groups, is exhibited in the Appendix. Hence we do have a basis for comparison of the two theories.

3.1. Theorems on free loops and their generalizations. We have the following general existence theorem:

THEOREM 3.1. (*General Existence Theorem*). *A half-loop H is contained in one, and essentially only one, loop T which is freely generated by H .*

Proof. By Theorem 1.2, $N(H)$ may be embedded in one and essentially only one net M , which is freely generated by $N(H)$. Clearly, $M = (M, O)$. Let $T = L(M, O)$. Then T is a loop which is freely generated by H , (Lemma 2.4). Furthermore, T is essentially uniquely determined, since the net construction is essentially unique.

The statement of Theorem 3.1 is totally false for half-groups or Abelian half-groups,⁸ since it is not true that every half-group (Abelian half-group)

⁸ An (Abelian) half-group is a half-loop which is associative (and commutative). A semi-group is an associative semi-loop.

may be embedded in a group (Abelian group). This is seen in the following discussion.

Malcev⁹ showed that, while a commutative semi-group⁸ is always embeddable in a group, there exist non-commutative semi-groups which can not be embedded in a group. We use Malcev's necessary condition for the embedding of a semi-group in a group to exhibit an example of an Abelian half-group which is not embeddable in a group:

Example. The following condition is easily verified to be necessary for the embedding of an Abelian half-group into an Abelian semi-group:

(A) If a, b, c, d, x, y, z, w , are any eight elements of an Abelian half-group satisfying the relations:

$$a + x = x + a = b + y = y + b; \quad c + x = x + c = d + y = y + d; \\ a + z = z + a = b + w = w + b;$$

then, necessarily, $c + z = z + c = d + w = w + d$.

Consider, then, the Abelian half-group H consisting of the fourteen distinct elements: $o, a, b, c, d, x, y, z, w, p, q, r, s, t$, with the following sums defined:

$$o + e = e + o = e, \text{ for every } e \text{ in } H; \\ a + x = x + a = b + y = y + b = p; \quad c + x = x + c = d + y = y + d = q \\ a + z = z + a = b + w = w + b = r; \quad c + z = z + c = s; \\ d + w = w + d = t.$$

Then H does not satisfy condition (A) and, hence, cannot be embedded in an Abelian semi-group (or group).

COROLLARY. *The half-loop K is a loop, if and only if, K has the following property:*

(I) If T, R , are loops generated by K , then there exists an isomorphism of T upon R which leaves every element of K fixed.

Proof. If K is a loop, then K enjoys property (I) trivially.

Suppose that K is a half-loop, but not a loop. Then K may be embedded in one and essentially only one loop T , which is freely generated by K (Theorem 3.1). By Theorem 1.3, $N(K)$ may be embedded in a net M , which is generated by $N(K)$, but not free over $N(K)$. Let $R = L(M, O)$;

⁹ See Malcev (1).

then R is a loop which is generated by K , but not free over K (Lemma 2.4). Now if there were an isomorphism between T and R , leaving every element of K fixed, then both T and R would be free over K , which is impossible. Hence K does not have property (I).

Note. This Corollary is false in the associative case, since Abelian semi-groups may be embedded in one and essentially only one smallest group, which is necessarily Abelian.

Theorem 3.1 proves the existence of free loops, so that we have immediately the following theorem, which is true also for groups¹⁰ and for Abelian groups.¹¹

THEOREM 3.2. *Every loop is the homomorphic image of a free loop.*

This theorem follows from the net theorem, Theorem 1.4, or may be verified directly.

Definition. The subloop L' of the loop L is the *commutator-associator*¹² subloop of L , if it is generated by all the elements x in L which satisfy at least one of the following conditions:

- | | |
|---|--------------------------------------|
| (i) $(a + b) + x = b + a$ | (iv) $[a + (b + c)] + x = (a + b) +$ |
| (ii) $x + (a + b) = b + a$ | (v) $x + [(a + b) + c] = a + (b +$ |
| (iii) $[(a + b) + c] + x = a + (b + c)$ | (vi) $x + [a + (b + c)] = (a + b) +$ |

for a, b, c , elements of L .

LEMMA 3.1. *If S, T , are two free sets of generators of the free loop L , and if n, m , are cardinal numbers representing the number of elements in S, T , respectively, then $n = m$.*

Note. We define, then, the *rank* of a free loop as the number of non-zero elements in a free set of generators of the loop.

Proof. Form the free Abelian group A on the generators S , and the free Abelian group B on the generators T .

Then the identity map of $S \subset L$ upon $S \subset A$ may be extended to a homomorphism η of L upon $L\eta = A$, since L is free over S . $K(\eta)$, the kernel of η , is the commutator-associator subloop L' of L . Hence $L/L' \sim A$.

¹⁰ See Kurosch (1).

¹¹ See Lefschetz (1), p. 50.

¹² The subloop L' has been introduced by Bruck ((2), pp. 267, 273) under the name of "the centrally derived loop of L ."

Similarly, the identity map of $T \subset L$ may be extended to a homomorphism β of L upon $L\beta = B$, and $K(\beta) = L'$, so that $L/L' \sim B$. Hence $A \sim B$; but isomorphic free Abelian groups have the same rank,¹³ and therefore, $n = m$.

COROLLARY. *If L, H , are free loops of rank n, m , respectively, then $L \sim H$, if and only if $n = m$.*

This is an immediate consequence of the Lemma.

THEOREM 3.3. *If the loop T is freely generated by its sub-half-loop $K \neq T$, then T contains a free subloop of infinite rank.*

Proof. $N(T)$ is free over and generated by its sub-half-net $N(K)$, and $N(K) \neq N(T)$ (Lemma 2.4, and property 1°, 2). By Theorem 1.7 $N(T)$ contains a free subnet S on a countably infinite free set of generators K . It is clear from the proof of Theorem 1.7 that we may require the point O to be in K . Then S has an O -centered free set of generators (K, O) , and $L(K, O) = R$ is therefore a free subloop of T having a countably infinite free set of generators, $L(K, O)$.

COROLLARY. *A free loop $L \neq 0$ contains a free subloop of countably infinite rank.*

Note that, if L is a free loop of finite rank, it is clear that the set L is countable, since a finite number of elements is introduced at each of the countable number of stages in the construction of L . Therefore the statement that L contains a free subloop of countable rank is a "best possible result."

Remark. By a "ladder" construction similar to the one used in Theorem 1.7, we obtain the following set of elements G which constitute a free set of generators for a subloop of infinite rank of the free loop on one generator, a :

$$G = \{a + a, a + [a + (a + a)], a + (a + \{a + [a + (a + a)]\}), \dots\}.$$

Corresponding to the above Corollary, we have the following theorems for groups and Abelian groups:

If H is a free group of rank greater than or equal to 2, then H contains a free subgroup of infinite rank.¹⁴ If A is a free Abelian group, and if B is a subgroup of A , then the rank of B cannot exceed the rank of A .¹⁵

LEMMA 3.2. *If $L = J * K$, and if M is the cross-cut of all normal subloops¹⁶ of L containing K , then $J \sim L/M$.*

¹³ See Lefschetz (1), p. 50.

¹⁴ See Levi (1).

¹⁵ See Lefschetz (1), p. 50.

¹⁶ See Baer (4).

Proof. Let ϕ be the homomorphism of J into L such that $j\phi = j$, for every j in J , and let ψ be the homomorphism of K into L , such that $K\psi = 0$. Since $L = J * K$, then ϕ and ψ may be extended simultaneously to a homomorphism η of L into L , i. e., to an endomorphism of L . Then $J = L\eta$, and $j\eta = j$, for every j in J .

We verify that the kernel of η is M , as follows: Suppose that ρ is any homomorphism of L such that $K\rho = 0$; then $L\rho$ is generated by $J\rho$, and hence $L\rho = J\rho$. Then $L\rho = J\rho = L\eta\rho$, and $\rho = \eta\rho$, since $(j)\eta\rho = (j\eta)\rho = j\rho$, for every j in J , (see Appendix, Theorem 4.1). This implies that the kernel of η is a part of the kernel of ρ . That is, the kernel of η is M . Hence we have $J \sim L/M$.

THEOREM 3.4. (Subloop Theorem). *If the loop R is freely generated by its half-loop T , and if H is a subloop of R , then H is the free sum of loops G and F , where*

i. G is freely generated by $H \circ T$.

ii. F is a free loop with a free set of generators

$$F_0 = \bigcup_{i=0}^{\infty} \{(H \circ T_{i+1}) - (H \circ T_i)_1\},$$

where $T \subset T_1 \subset \dots \subset T_i \subset \dots \subset R$, is the generating chain of T in R .

Proof. $N(R)$ is freely generated by $N(T)$ (Lemma 2.4). By Theorem 1.5, the subnet $N(H)$ of $N(R)$ is the free sum of nets S and W , where S is a free net with a free set of generators

$$S_0 = \bigcup_{j=1}^{\infty} \{[N(H) \circ N(T)^{(2j+1)}] - [N(H) \circ N(T)^{(2j)}]_L\},$$

and W is the subnet of $N(H)$ which is freely generated by

$$N(H) \circ N(T) = N(H \circ T).$$

Now $N(H)$, $N(T)$, $N(H \circ T)$, are all O -centered; then $N(H)$ is the O -centered free sum of the nets W and S^* , where S^* is generated by (S_0, O) , the smallest O -centered half-net containing S_0 and the point O (Theorem 1.10).

Since $W = (W, O)$ is freely generated by $N(H \circ T)$, then $G = L(W, O)$ is freely generated by $H \circ T$ (Lemma 2.4).

S^* is a free net with an O -centered free set of generators (S_0, O) (Theorem 1.9), and hence $F = L(S^*, O)$ is a free loop with a free set of generators, $L(S_0, O)$ (Property 5°, 2).

Hence $H = F * G$, and we have only to show that $F_0 = L(S_0, O)$:

Consider any element which is in F_0 , and hence is in the half-loop, $(H \circ T_{i+1}) - (H \circ T_i)_1$, for some integer i ; there are three possibilities to consider:

(1) z in $(H \circ T_{i+1}) - (H \circ T_i)_1$ (and hence z in T_{i+1} , but not in T_i), such that $z = a + b$; a, b , in T_i .

By Lemma 2.5, there is an integer $j \geq 0$, such that $r(a)$ and $s(b)$ are in $N(T)^{(2j+1)}$, and such that (a, b) is not in $N(T)^{(2j+1)}$. Then $(a, b) = r(a) \cdot s(b)$ is in $N(T)^{(2j+2)}$, but not in $N(T)^{(2j+1)}$, and $r(a), s(b)$, are the only lines on (a, b) in $N(T)^{(2j+2)}$. Hence $t(z) = t(a, b)$ is in $N(T)^{(2j+3)}$, but not in $N(T)^{(2j+2)}$, and $t(z)$ carries (a, b) and no other point in $N(T)^{(2j+3)}$.

Since z is in $(H \circ T_{i+1}) - (H \circ T_i)_1$, then z is in H , but neither a nor b is in H (for, if both a, b , were in H , then $z = a + b$ would be in $(H \circ T_i)_1$, which is impossible; and if one of the elements a, b were in H , then the other would also be in H , since z is in H).

Then $r(a)$ is not in $N(H)$, since, if it were, the point (a, o) on $r(a)$ and $s(o)$ would be in $N(H)$ and a would be in H . Similarly, the line $s(b)$ is not in $N(H)$. Consequently, the point (a, b) is not in $N(H)$. However, the line $t(z) = t(z, o)$ is in $N(H)$, since z is in H .

Thus we have that $t(z)$ is in

$$[N(H) \circ N(T)^{(2j+3)}] - [N(H) \circ N(T)^{(2j+2)}]_L,$$

and therefore, in S_0 . Then $(z, o) = t(z) \cdot s(o)$ is in (S_0, O) , and z is in $L(S_0, O)$.

(2) x is in $(H \circ T_{i+1}) - (H \circ T_i)_1$, such that $x + a = b$; a, b , in T_i .

By Lemma 2.5, there exists an integer $k \geq 0$, such that $s(a)$ and $t(b)$ are both in $N(T)^{(2k+1)}$, and such that the point (x, a) is not in $N(T)^{(2k+1)}$. Then $(x, a) = s(a) \cdot t(b)$ is in $N(T)^{(2k+2)}$, but not in $N(T)^{(2k+1)}$, and $s(a), t(b)$, are the only lines on (x, a) in $N(T)^{(2k+2)}$. Hence $r(x, a) = r(x)$ is in $N(T)^{(2k+3)}$, but not in $N(T)^{(2k+2)}$, and the only point on $r(x)$ in $N(T)^{(2k+3)}$ is (x, a) .

Since x is in $(H \circ T_{i+1}) - (H \circ T_i)_1$, then x is in H , but neither a nor b is in H . Then neither $s(a)$ nor $t(b)$ is in $N(H)$, although $r(x)$ is in $N(H)$. That is, $r(x)$ is in $[N(H) \circ N(T)^{(2k+3)}] - [N(H) \circ N(T)^{(2k+2)}]_L \subset S_0$. Hence $(x, o) = r(x) \cdot s(o)$ is in (S_0, O) and x is in $L(S_0, O)$.

(3) y in $(H \circ T_{i+1}) - (H \circ T_i)_1$, such that $a + y = b$, for a, b , in T_i .

By Lemma 2.5, there exists an integer $l \geq 0$, such that $r(a)$ and $t(b)$

are in $N(T)^{(2l+1)}$, and such that the point (a, y) is not in $N(T)^{(2l+1)}$. Then $(a, y) = r(a) \cdot t(b)$ is in $N(T)^{(2l+2)}$, but not in $N(T)^{(2l+1)}$, and $r(a)$, $t(b)$, are the only lines on (a, y) in $N(T)^{(2l+2)}$. Hence $s(y) = s(a, y)$ is in $N(T)^{(2l+3)}$, but not in $N(T)^{(2l+2)}$, and (a, y) is the only point on $s(y)$ in $N(T)^{(2l+3)}$.

Since y is in $(H \circ T_{i+1}) - (H \circ T_i)_1$, then y is in H , but neither a nor b is in H . Then neither $r(a)$ nor $t(b)$ is in $N(H)$, although $s(y)$ is in $N(H)$. That is, $s(y)$ is in $[N(H) \circ N(T)^{(2l+3)}] - [N(H) \circ N(T)^{(2l)}]_L \subset S_0$. Then $(o, y) = s(y) \cdot r(o)$ is in (S_0, O) and y is in $L(S_0, O)$.

This completes our proof showing that $F_0 \subset L(S_0, O)$.

We verify, next, that H is generated by

$$(H \circ T) \cup F_0 = (H \circ T) \cup \bigcup_{i=0}^{\infty} \{(H \circ T_{i+1}) - (H \circ T_i)_1\}.$$

For, we may represent H in the form: $H = \bigcup_{i=0}^{\infty} (H \circ T_i)$, since $H \subset R = \bigcup_{i=0}^{\infty} T_i$.

But if we let $K(n) = (H \circ T_n) \cup \bigcup_{i=n}^{\infty} \{(H \circ T_{i+1}) - (H \circ T_i)_1\}$, then it is easily checked that H may also be represented in the form: $H = \bigcup_{i=0}^{\infty} K(i)$, and that H is generated by $K(o) = (H \circ T) \cup F_0$.

There exists one and only one endomorphism ϵ of H which leaves invariant every element in $(H \circ T) \cup F_0$, and which annihilates every element in $(L(S_0, O) - F_0)$ (since L is freely generated by $(H \circ T)$ and $L(S_0, O)$; and F_0 is a subset of the free set of generators, $L(S_0, O)$). But $\epsilon = 1$, since H is generated by $(H \circ T) \cup F_0$. Hence $(L(S_0, O) - F_0)$ is vacuous; i. e., $L(S_0, O) = F_0$.

We incorporate the general results of Theorem 3.4, together with some obvious inferences of this theorem into the following theorem for later convenience:

General Subloop Theorem. If the loop R is freely generated by its half-loop T , if H is a subloop of R , and if G is the subloop of H which is (freely) generated by $H \circ T$, then $H = F * G$, where:

i. F is a free loop whose rank is uniquely determined by H and T .

ii. F possesses a free basis F_0 , such that the following property is true:

(α) If α is an automorphism of R such that $H\alpha = H$ and $T\alpha = T$, then $F\alpha = F$ and α effects a permutation of the elements of F_0 .

Property i. of F is an immediate consequence of Lemma 3.2, and

property ii. is obviously true of the particular choice of the free loop F and basis F_0 in Theorem 3.4.

COROLLARY. *A subloop of a free loop is free.*

Proof. If the loop R is free, then R has a free set of generators, J . If H is a subloop of R , then, by the theorem above, $H = F * G$, where F is a free loop and where G is the subloop of H that is freely generated by $H \cap J$. Since J is a free set of generators, then it is obvious that $H \cap J$ forms a free set of generators of G , and hence that G is a free loop. But the free sum of free loops is free; hence H is a free loop.

The corresponding statements to the Corollary for groups,¹⁷ and for Abelian groups¹⁸ are true.

THEOREM 3.5. *K is a free loop, if and only if, for every loop H and homomorphism ϕ such that $H\phi = K$, there exists a (free) subloop S of H , such that ϕ induces an isomorphism of S upon K .*

This theorem may be deduced readily from Theorem 1.6 by the customary arguments involving net representations or may be verified directly by a proof similar to the one of Theorem 1.6.

Theorem 3.5 is true also for groups¹⁹ and for Abelian groups.²⁰

3.2. Theorems on free sums and generalized free sums of loops.

We prove first the following lemma:

LEMMA 3.3. *Given $\{L(v)\}$, a set of loops having the same 0-element, then $S = \bigcup L(v)$ is a half-loop, if S has the property:*

(A) *If two of the elements a, b, c , are in $L(v) \cap L(w)$, then $a + b = c$ in $L(v)$, if and only if $a + b = c$ in $L(w)$.*

Proof. The expression $a + b = c$, for a, b, c , in S , has meaning only if a, b, c , are all in some $L(v)$, with addition as defined in that $L(v)$. Hence if $a + b = c$, and $a + b = d$, in S , then $c = d$, by (A); and if $a + f = a + g$ in S , then $f = g$; or if $h + a = k + a$ in S , then $h = k$. Hence S is a half-loop.

We now generalize the concept of "free sum" in the following way:

¹⁷ See Schreier (1).

¹⁸ See Lefschetz (1), p. 50.

¹⁹ See Baer (5).

²⁰ See Eilenberg-MacLane (1).

Definition. The loop L is the *generalized free sum* of its subloops $L(v)$, if the following conditions are satisfied:

- (a) L is generated by $S = \bigcup L(v)$.
- (b) If $\alpha(v)$ are homomorphisms of the $L(v)$, with the property:
 - (*) If x is in $L(v) \circ L(w)$, then $x\alpha(v) = x\alpha(w)$,

then there exists a homomorphism α of L , which induces $\alpha(v)$ in $L(v)$.

Note. If x in $L(v) \circ L(w)$, $v \neq w$, implies that $x = 0$, then L is simply the *free sum* of its subloops $L(v)$.

Remark. If the half-loop H is the join of its sub-half-loops, $H(v)$, then every homomorphism α of H induces well-determined homomorphisms $\alpha(v)$ of $H(v)$ with the property

$$(*) \text{ If } x \text{ is in } H(v) \circ H(w), \text{ then } x\alpha(v) = x\alpha(w).$$

If, conversely, there is given for every v , a homomorphism $\alpha(v)$ of $H(v)$, such that $(*)$ is valid, then there exists one and only one homomorphism α of H inducing $\alpha(v)$ in $H(v)$. (Note that this last statement is no longer true if we have, instead, that H is generated by the $H(v)$, since the existence of the homomorphism of H would imply freeness.)

By virtue of the statements in the above remark, we may give the equivalent definition:

Definition. The loop L is the *generalized free sum* of its subloops $L(v)$, if and only if, L is freely generated by the join of the $L(v)$.

Note. If L is the (generalized) free sum of its subloops $L(v)$, then L is the (generalized) free sum of $L(w)$ and T , where T is the (generalized) free sum of the loops $L(v)$, for $v \neq w$.

If $\{L(v)\}$ is a set of loops, having the same o -element, and satisfying condition (A) of Lemma 3.3, then it is clear that L , the generalized free sum of the $L(v)$ exists. For, $S = \bigcup L(v)$ is a half-loop and consequently may be embedded in a loop L which is freely generated by S (Theorem 3.1).

Suppose that there is given a set of loops $\{L(v)\}$, and that there is given for every pair of indices $v \neq w$ an isomorphic correspondence $\alpha(v, w)$ mapping a certain subloop $L(v; w)$ of $L(v)$ upon a subloop $L(v; w)\alpha(v, w)$ of $L(w)$; these isomorphisms are subject to the following requirements:

- (a) $\alpha(v, w) = \alpha(w, v)^{-1}$, so that $L(v; w)\alpha(v, w) = L(w; v)$.
- (b) If x is an element in $L(v; w)$, and if $x\alpha(v, w)$ belongs to $L(w; z)$,

(so that $x\alpha(v, w)$ is in $L(w; v) \circ L(w; z)$), then x is in $L(v; z)$ and $x\alpha(v, z) = x\alpha(v, w)\alpha(w, z)$.

Then it is possible to construct a half-loop H which is the join of the $L(v)$ with the element x in $L(v; w)$ identified with the element $x\alpha(v, w)$ in $L(w; v)$, because now condition (A) of Lemma 3.3 is satisfied. The loop L which is freely generated by this half-loop H is called the *free sum of the $L(v)$ with amalgamated subloops*.

We use the terms "generalized free sum" and "free sum with amalgamated subloops" interchangeably—it being understood that the identification of corresponding elements has already been effected.

The existence of the generalized free sum (product) of groups, in the sense of our definition, seems to be an open question. Schreier²¹ has proved the existence of the free sum of groups $G(v)$ with amalgamated subgroups $H(v)$, for the special case in which each $H(v)$ is isomorphic to a given group H . The existence of direct sums of Abelian groups with amalgamated subgroups also appears to be not a trivial question.

We shall use the notation: $L = \sum^* L(v)$ to denote the *generalized free sum of the $L(v)$* .

THEOREM 3.6. (*Subloop Theorem for Subloops of a (Generalized) Free Sum*) If $L = \sum^* L(v)$ and if H is a subloop of L , then $H = F * \sum^* [H \circ L(v)]$, where F is a free loop such that:

- (i) The rank of F is uniquely determined by the $L(v)$ and H .
- (ii) There exists a free basis F_0 of F such that if α is an automorphism of L with $L(v)\alpha = L(v)$, for every v , and $H\alpha = H$, then $F\alpha = F$ and α effects a permutation of the elements in F_0 .

Proof. Since L is the generalized free sum of its subloops $L(v)$, then L is freely generated by its sub-half-loop $S = \bigcup L(v)$ (where suitable identification of elements in the $L(v)$ have been made). By the General Subloop Theorem, then H is the free sum of loops F and G , where G is the subloop of H which is freely generated by $H \circ S$, and where F is a free loop with the following properties:

- (i) The rank of F is uniquely determined by H and S , and therefore, by H and the $L(v)$.
- (ii) There exists a free basis F_0 of F , such that if α is an automorphism of L with $H\alpha = H$ and $S\alpha = S$, then $F\alpha = F$, and α effects a permutation of

²¹ See Schreier (1).

the elements of F_0 . Now if $L(v)\alpha = L(v)$, then certainly $S\alpha = S = \bigcup L(v)$, so that we have property (ii) as stated in this theorem.

Since $H \circ S = H \circ \bigcup L(v) = \bigcup [H \circ L(v)]$ and G is freely generated by $H \circ S$, then G is the generalized free sum of the $H \circ L(v)$, which concludes the proof.

COROLLARY. If $L = \sum^* L(v)$, and if H is a subloop of L , then $H = F * \sum^* [H \circ L(v)]$.

The corresponding theorem for subgroups of a free product of groups is as follows:²²

If the group G is the free product of its subgroups $G(v)$, and if U is a subgroup of G , then $U = F * \prod^* U(i, v)$, where

i. F is a free group, and $U(i, v) = U \circ u_i G(v) u_i^{-1}$, with $v = v(i)$ and u_i in G .

ii. There is one and only one $U(j, v)$ in U which is conjugate to $U \circ w G(v) w^{-1} \neq 1$, for arbitrary v, w .

That our Corollary is not true for groups is exhibited in the following example:

Example. Consider $G = G_1 * G_2$, where G_1 is a cyclic group generated by a , and G_2 is a finite cyclic group generated by b . Then H , the subgroup of G which is generated by $a^{-1}ba$, does not have the structure required in the Corollary, since H is not free, although its crosscuts with G_1 and G_2 are zero.

There is no corresponding statement to the Corollary which can be made for subgroups of a direct sum of Abelian groups. For example, there exist Abelian groups without elements of finite order (except o), which are not direct sums of groups of rank 1.²³ Such a group, however, is a subgroup of an Abelian group which is the direct sum of groups, each of rank 1.

THEOREM 3.7. If $L = \sum^* L_i = \sum^* K_j$ are two decompositions of the loop L into a free sum of loops, then there exist free loops F_i, G_j , such that

$$L_i = F_i * \sum_j^* (L_i \circ K_j); \quad K_j = G_j * \sum_i^* (L_i \circ K_j),$$

and

$$L = \sum_{i,j}^* F_i * \sum_{i,j}^* (L_i \circ K_j) = \sum_{i,j}^* G_j * \sum_{i,j}^* (L_i \circ K_j),$$

where $\sum^* F_i$ and $\sum^* G_j$ are isomorphic free loops.

²² See Baer and Levi (1) and Takahasi (1).

²³ See Baer (2)

Proof. This theorem is an immediate consequence of Theorem 3.4 and Lemma 3.2.

The corresponding theorem for the decomposition of groups into free products is the following:²⁴

If $G = \prod_v^* A_v = \prod_w^* B_w$ are two decompositions of the group G into a free product of groups, then there are free groups F_v and G_w , such that

$$A_v = \prod_k^* A_{vk} * F_v \text{ and } B_w = \prod_i^* B_{wi} * G_w, \text{ and}$$

$$G = \prod_{v,k}^* A_{vk} * \prod_v^* F_v = \prod_{w,i}^* B_{wi} * \prod_w^* G_w,$$

where A_{vk} and B_{wi} are pairwise conjugate and $\prod_v^* F_v$, $\prod_w^* G_w$ are isomorphic free groups.

There is no corresponding theorem for direct sums of Abelian groups.²⁵

APPENDIX.

We consider a class \mathfrak{R} of half-loops meeting the following requirements:

- (I) If S is a sub-half-loop of the half-loop H in \mathfrak{R} , then S is in \mathfrak{R} .
- (II) If the half-loop J is the homomorphic image of the half-loop H in \mathfrak{R} , then J is in \mathfrak{R} .

It is clear that the following three classes meet requirements (I) and (II) for a class \mathfrak{R} : the class \mathfrak{C} of all half-loops; the class \mathfrak{G} of all associative half-loops (i. e. half-groups); and the class \mathfrak{A} of all associative and commutative half-loops, (i. e. Abelian half-groups). While these three classes by no means exhaust the possibilities, they are the only ones with which we shall here be concerned.

Definition. The loop L in \mathfrak{R} is generated by its sub-half-loop J if no proper subloop of L contains J .

THEOREM 4.1. *If the loop L in \mathfrak{R} is generated by its sub-half-loop J , and if ρ, σ , are homomorphisms of L into the loop S in \mathfrak{R} , such that $j\rho = j\sigma$, for every j in J , then $\rho = \sigma$.*

Proof. Let T be the set of elements t in L , such that $t\rho = t\sigma$, for every

²⁴ See Baer and Levi (1) and Takahasi (1).

²⁵ See Baer (1).

t in T . Clearly T contains J , and, since T is a sub-half-loop of L , T is in \mathfrak{R} . But then it is easily verified that T is, moreover, a subloop of L , and hence, since L is generated by J , we have $T = L$. That is $\rho = \sigma$.

Definition. The half-loop L in \mathfrak{R} is \mathfrak{R} -free over its sub-half-loop J if it meets the following requirement:

(F) If σ is a homomorphism of J into the loop S in \mathfrak{R} , then there exists a homomorphism τ of L into the same loop S , which coincides with σ in J . (Note that $J\sigma$ and $L\tau$ are necessarily in \mathfrak{R} .)

THEOREM 4.2. *Given a half-loop J in \mathfrak{R} , then there is essentially at most one loop L in \mathfrak{R} , which is \mathfrak{R} -freely generated by J .*

Proof. Suppose that L and K are loops in \mathfrak{R} , each of which is \mathfrak{R} -freely generated by J . Since L is \mathfrak{R} -free over J , the identity map of $J \subset L$ upon $J \subset K$ may be extended to a homomorphism η of L into K ; since K is generated by J , η is a homomorphism of L upon K , and we have $L\eta = K$. But we may extend, similarly, the identity map of $J \subset K$ upon $J \subset L$ to a homomorphism τ of K upon L , obtaining, then $K\tau = L$. Hence $L\eta\tau = L$, and $j\eta\tau = j$, for every j in J . By Theorem 4.1, then $\eta\tau = 1$. Similarly, $\tau\eta = 1$. Hence η and τ are reciprocal isomorphisms of L and K , and we have $L \sim K$.

Of particular interest are the \mathfrak{R} -free loops and the \mathfrak{R} -free sums of loops, which we define now:

Definition. If the loop L in \mathfrak{R} is \mathfrak{R} -free over and generated by its sub-half-loop J , and if $x + y = z$ in J if and only if x or y is o , then we define J to be a \mathfrak{R} -free set of generators of L . A loop L in \mathfrak{R} is \mathfrak{R} -free if it possesses a \mathfrak{R} -free set of generators.

Note that if J is a \mathfrak{R} -free set of generators of the \mathfrak{R} -free loop L , then every single-valued mapping η of J into the loop M in \mathfrak{R} , which maps the null element of J upon the null element of M , may be extended to a homomorphism of L into M .

Definition. The loop L in \mathfrak{R} is the generalized \mathfrak{R} -free sum of its subloops $L(i)$, if it meets the following requirements:

(i) L is generated by $\bigcup L(i)$.

(ii) If there is given, for every i , a homomorphism $\alpha(i)$ of $L(i)$ (into the loop S in \mathfrak{R}), such that $x\alpha(i) = x\alpha(j)$, for every x in $L(i) \cap L(j)$,

then there exists a homomorphism α of L into S which coincides with $\alpha(i)$ in $L(i)$.

Note. Conditions (i) and (ii) are equivalent to requiring that L be \mathfrak{R} -freely generated by the set-theoretical sum of its subloops $L(i)$, which sum is a sub-half-loop of L .

If $L(i) \circ L(j) = o$, for every $i \neq j$, and if L meets (i) and (ii) above, then L is termed the \mathfrak{R} -free sum of its subloops $L(i)$.

Note, in particular, that a \mathfrak{R} -free loop is the \mathfrak{R} -free sum of \mathfrak{R} -free cyclic loops.

Now it is easy to verify that an \mathfrak{M} -free loop is a free Abelian group, a \mathfrak{G} -free loop is a free group, an \mathfrak{M} -free sum of loops is a direct sum of Abelian groups, and a \mathfrak{G} -free sum is a free sum of groups. For example, suppose that L is an \mathfrak{M} -free loop; then L has an \mathfrak{M} -free set of generators J . Construct the free Abelian group L^* on the set of generators J . Then every element in L^* has a unique representation in terms of the elements of J , and hence, given a single-valued map ρ of J into an Abelian group T , ρ may be extended to a homomorphism σ of L^* into T in the usual manner (i. e., for x in L^* , then $x = \sum n_i j_i$, where $j_i \in J$, and the n_i are integers; define $x\sigma = \sum n_i (j_i)\rho$). Hence L^* is an \mathfrak{M} -free loop with J as an \mathfrak{M} -free set of generators. But then $L^* \sim L$, since a \mathfrak{R} -free set of generators essentially uniquely determines a \mathfrak{R} -free loop (Theorem 4.2).

UNIVERSITY OF ILLINOIS,
URBANA, ILLINOIS.

BIBLIOGRAPHY

A. A. Albert

- (1) "Quasigroups I," *Transactions of the American Mathematical Society*, vol. 54 (1943), pp. 507-519.
- (2) "Quasigroups II," *Transactions of the American Mathematical Society*, vol. 55 (1944), pp. 401-419.

R. Baer

- (1) "The Decomposition of Enumerable, Primary, Abelian Groups into Direct Summands," *Quarterly Journal of Mathematics*, Oxford Series, vol. 6 (1935), pp. 217-221.
- (2) "Abelian groups without elements of finite order," *Duke Mathematical Journal*, vol. 3 (1937), pp. 68-122.
- (3) "Nets and groups I," *Transactions of the American Mathematical Society*, vol. 46 (1939), pp. 110-141.

- (4) "The homomorphism theorems for loops," *American Journal of Mathematics*, vol. 47 (1945), pp. 450-460.
 - (5) "Representations of groups as quotient groups I," *Transactions of the American Mathematical Society*, vol. 58 (1945), pp. 295-347.
- R. Baer and F. Levi
- (1) "Freie Produkte und ihre Untergruppen," *Compositio Mathematica*, vol. 3 (1936), pp. 391-398.
- G. Bol
- (1) "Gewebe und Gruppen," *Mathematische Annalen*, vol. 114 (1937), pp. 414-431.
- R. Bruck
- (1) "Simple quasigroups," *Bulletin of the American Mathematical Society*, vol. 50 (1944), pp. 769-781.
 - (2) "Contributions to the theory of loops," *Transactions of the American Mathematical Society*, vol. 60 (1946), pp. 245-354.
- E. Eilenberg—S. MacLane
- (1) "Group extensions and homologies," *Annals of Mathematics*, vol. 43 (1942), pp. 757-831.
- M. Hall
- (1) "Projective planes," *Transactions of the American Mathematical Society*, vol. 54 (1943), pp. 229-277.
- A. Kurosch
- (1) "Über freie Produkte von Gruppen," *Mathematische Annalen*, vol. 108 (1933), pp. 26-36.
- S. Lefschetz
- (1) "Algebraic topology," *American Mathematical Colloquium Publications*, vol. 27, New York (1942).
- F. Levi
- (1) "Über die Untergruppen freier Gruppen," *Mathematische Zeitschrift*, vol. 32 (1930), pp. 315-318.
- A. Malcev
- (1) "On the immersion of an algebraic ring into a field," *Mathematische Annalen*, vol. 113 (1937), pp. 686-691.
- O. Schreier
- (1) "Die Untergruppen der freien Gruppen," *Hamburger Abhandlungen*, vol. 5 (1927), pp. 161-183.
- M. F. Smiley
- (1) "An application of lattice theory to Quasigroups," *Bulletin of the American Mathematical Society*, vol. 50 (1944), pp. 782-786.
- M. Takahasi
- (1) "Bemerkungen über den Untergruppensatz in freien Produkten," *Proceedings of the Imperial Academy*, Tokyo, vol. 20 (1944), pp. 589-594.

ON THE CONVERGENCE AND DIVERGENCE OF CONTINUED FRACTIONS.*

By W. T. SCOTT and H. S. WALL.

1. Introduction. A sequence x_1, x_2, x_3, \dots of complex numbers is said to satisfy *condition (H)* if at least one of the following statements holds.

- (a). The series $\sum |x_{2p+1}|$ diverges.
- (b). The series $\sum |x_{2p+1}(x_2 + x_4 + \dots + x_{2p})^2|$ diverges.
- (c). $\lim_{p \rightarrow \infty} |x_2 + x_4 + \dots + x_{2p}| = \infty$.

This condition was first used in connection with continued fractions by Hamburger [2]. In this paper we show that a necessary condition for the continued fraction $\sum_{p=1}^{\infty} (1/b_p)$ to converge is that the sequence $\{b_p\}$ of its partial denominators satisfy condition (H). In particular, the continued fraction diverges if the series $\sum |b_p|$ converges (Stern [6], von Koch [4]), or, more generally, if the two series $\sum b_{2p}$ and $\sum b_{2p+1}$ converge, at least one absolutely. We show further that if $b_{2p-1} = k_{2p-1}z_p$, $b_{2p} = k_{2p}$, $k_1 > 0$, $k_{2p-1} \geq 0$, $R(k_{2p}) \geq 0$, $R(z_p) \geq \delta$, $|z_p| \leq M$, $p = 1, 2, 3, \dots$, where δ and M are positive constants, then the continued fraction converges if and only if the sequence $\{k_p\}$ satisfies condition (H). This result includes and correlates convergence theorems of Stieltjes [7], Van Vleck [10], Hamburger [2], and Mall [5].

2. Necessity of condition (H). We shall now prove the following theorem.

THEOREM A. *A necessary condition for the continued fraction $\sum_{p=1}^{\infty} (1/b_p)$ to converge is that the sequence $\{b_p\}$ of its partial denominators satisfy condition (H).*

Proof. We suppose that the sequence $\{b_p\}$ does not satisfy condition (H), i. e., that the series

$$(2.1) \quad \sum |b_{2p+1}|, \quad \sum |b_{2p+1}(b_2 + b_4 + \dots + b_{2p})^2|$$

* Received August 20, 1946; Revised April 26, 1947.

converge, and that the sequence $s_p = b_2 + b_4 + \cdots + b_{2p}$, $p = 1, 2, 3, \cdots$, has at least one finite limit-point. Then the series $\sum |b_{2p+1}s_p|$ converges, and there exists an index n such that $|b_{2p+1}s_p| < 1$ for $p > n$, so that the infinite product $\prod_{p=1}^{\infty} (1 + b_{2n+2p+1}s_{n+p})$ converges and is not zero.

Let A_p and B_p denote the p -th numerator and denominator of the continued fraction. Then

$$(2.2) \quad A_{p+1} = b_{p+1}A_p + A_{p-1}, \quad B_{p+1} = b_{p+1}B_p + B_{p-1}, \quad p = 0, 1, 2, \cdots,$$

where $A_{-1} = 1$, $B_{-1} = 0$, $A_0 = 0$, $B_0 = 1$. Put

$$(2.3) \quad \begin{aligned} \pi_0 &= 1, \quad \pi_p = \prod_{r=1}^p (1 + b_{2n+2r+1}s_{n+r}), & p &= 1, 2, 3, \cdots, \\ U_{2p} &= A_{2n+2p+1}/\pi_p, & V_{2p} &= B_{2n+2p+1}/\pi_p, \\ U_{2p+1} &= (A_{2n+2p+2} - s_{n+p+1}A_{2n+2p+1})\pi_p, \\ V_{2p+1} &= (B_{2n+2p+2} - s_{n+p+1}B_{2n+2p+1})\pi_p, \\ c_{2p} &= b_{2n+2p+1}/\pi_{p-1}\pi_p, & c_{2p+1} &= -b_{2n+2p+1}s_{n+p}^2\pi_p\pi_{p-1}. \end{aligned}$$

One may then readily verify that

$$(2.4) \quad U_p = c_p U_{p-1} + U_{p-2}, \quad V_p = c_p V_{p-1} + V_{p-2}, \quad p = 2, 3, 4, \cdots.$$

Inasmuch as the series (2.1) are convergent and $\lim_{p \rightarrow \infty} \pi_p = \rho$ exists and is finite and not zero, it follows that the series $\sum |c_p|$ is convergent. Hence we conclude by a known argument (cf., for instance, Hellinger and Wall [3], p. 113) that the sequences $\{U_{2p}\}$, $\{V_{2p}\}$, $\{U_{2p+1}\}$, $\{V_{2p+1}\}$ have finite limits X_0 , Y_0 , X_1 , Y_1 , respectively. Therefore, by (2.3),

$$(2.5) \quad \begin{aligned} \lim_{p \rightarrow \infty} A_{2p+1} &= \rho X_0, & \lim_{p \rightarrow \infty} (A_{2p} - s_p A_{2p-1}) &= X_1/\rho, \\ \lim_{p \rightarrow \infty} B_{2p+1} &= \rho Y_0, & \lim_{p \rightarrow \infty} (B_{2p} - s_p B_{2p-1}) &= Y_1/\rho. \end{aligned}$$

Inasmuch as $A_{2p-1}(B_{2p} - s_p B_{2p-1}) - B_{2p-1}(A_{2p} - s_p A_{2p-1}) = 1$, it follows that

$$(2.6) \quad X_0 Y_1 - X_1 Y_0 = 1.$$

Let s be a finite limit-point of the sequence $\{s_p\}$, and suppose that $s_p \rightarrow s$ as $p \rightarrow \infty$ over a sequence P of indices. We then conclude from (2.5) that

$$(2.7) \quad \begin{aligned} \lim_{p \rightarrow \infty} A_{2p} &= s\rho X_0 + X_1/\rho = F(s), \\ \lim_{p \rightarrow \infty} B_{2p} &= s\rho Y_0 + Y_1/\rho = G(s), \end{aligned}$$

as $p \rightarrow \infty$ over P . Moreover, by (2.6),

$$X_0 G(s) - Y_0 F(s) = 1,$$

and

$$F(s) G(t) - F(t) G(s) = s - t.$$

If condition (H) is not satisfied by the sequence $\{b_p\}$, we now conclude immediately that the sequence of approximants of the continued fraction has distinct limit-points X_0/Y_0 , $F(s)/G(s)$, where the range of s is the set of finite limit-points of the sequence $\{s_p\}$. Hence condition (H) is necessary for the convergence of the continued fraction.

On applying the preceding considerations to the continued fraction $\sum_{p=2}^{\infty} (1/b_p)$, we conclude that $\sum_{p=1}^{\infty} (1/b_p)$ diverges if the series $\sum |b_{2p+2}|$ and $\sum |b_{2p+2}(b_3 + b_5 + \cdots + b_{2p+1})^2|$ converge and the sequence $b_3 + b_5 + \cdots + b_{2p+1}$, $p = 1, 2, 3, \cdots$, has a finite limit-point. In particular, we have the following extension of a theorem of von Koch [4].

THEOREM B. *The continued fraction $\sum_{p=1}^{\infty} (1/b_p)$ diverges if the series $\sum b_{2p}$ and $\sum b_{2p+1}$ converge, at least one of them absolutely. When the condition is satisfied, the sequences $\{A_{2p}\}$, $\{B_{2p}\}$, $\{A_{2p+1}\}$, $\{B_{2p+1}\}$ of even and odd numerators and denominators have finite limits F_0 , G_0 , F_1 , G_1 , respectively, where $F_1 G_0 - F_0 G_1 = 1$.*

3. The question of sufficiency of condition (H). The sequence $b_p = i$, $p = 1, 2, 3, \cdots$, satisfies condition (H) although the continued fraction $\sum_{p=1}^{\infty} (1/b_p)$ diverges in this case. Hence condition (H) is not, in general, sufficient for the convergence of the continued fraction. This is shown also by the fact that the continued fraction diverges if all its odd partial denominators b_{2p-1} are zero. In this case, the series (2.1) are both convergent, and condition (H) therefore reduces to the condition

$$(3.1) \quad \lim_{p \rightarrow \infty} |b_2 + b_4 + \cdots + b_{2p}| = \infty.$$

The following theorem covers a class of convergent continued fractions for which condition (H) takes this form.

THEOREM C. *If $R(b_1) > 0$, $R(b_p) \geq 0$, $p = 2, 3, 4, \cdots$, and if the series (2.1) are both convergent, then the continued fraction $\sum_{p=1}^{\infty} (1/b_p)$ converges if and only if (3.1) holds.*

Proof. The fact that (3.1) is necessary for convergence of the continued fraction is contained in Theorem A. Moreover, from (2.5) we see that when (3.1) holds, then

$$\lim_{p \rightarrow \infty} \frac{A_p}{B_p} = \frac{X_0}{Y_0}.$$

The finite numbers X_0 and Y_0 are not both zero, by virtue of (2.6). Since $R(b_p) \geq 0$, the linear transformation $t = t_p(w) = 1/(b_p + w)$ of the w -plane into the t -plane maps $R(w) \geq 0$ into $R(t) \geq 0$; since $R(b_1) > 0$, $t = t_1(w)$ maps $R(w) \geq 0$ upon the circular region

$$\left| t - \frac{1}{2R(b_1)} \right| \leq \frac{1}{2R(b_1)}.$$

It follows that the transformation $t = t_1 t_2 \cdots t_p(w)$ maps $R(w) \geq 0$ into this circular region. Therefore, since $t_1 t_2 \cdots t_p(0) = A_p/B_p$, we conclude that

$$(3.2) \quad \left| \frac{A_p}{B_p} \right| \leq \frac{1}{R(b_1)}, \quad p = 1, 2, 3, \cdots$$

Consequently, $Y_0 \neq 0$, so that the continued fraction converges to the value X_0/Y_0 .

Remark. It is easy to see that the conclusion in Theorem C holds if $b_1 = b_3 = \cdots = b_{2m-1} = 0$, $R(b_{2m+1}) > 0$, $R(b_p) \geq 0$ for $p > 2m + 1$, the series (2.1) are both convergent, and $\lim s_p = \infty$.

If at least one of the series (2.1) is divergent, we can establish convergence of the continued fraction $\bar{K}_{p=1}^{\infty}(1/b_p)$ under somewhat more restrictive hypotheses upon the b_p than were used in Theorem C.

THEOREM D. Let $b_{2p-1} = k_{2p-1}z_p$, $b_{2p} = k_{2p}$, $p = 1, 2, 3, \cdots$, where

$$(3.3) \quad k_1 > 0, k_{2p+1} \geq 0, R(k_{2p}) \geq 0, R(z_p) \geq \delta, |z_p| \leq M, p = 1, 2, 3, \cdots,$$

δ and M being positive constants. The continued fraction $\bar{K}_{p=1}^{\infty}(1/b_p)$ converges if and only if the sequence $\{k_p\}$ satisfies condition (H).

Remark. The continued fraction $\bar{K}_{p=1}^{\infty}(1/b_p)$ is equivalent to the continued fraction $i \cdot \bar{K}_{p=1}^{\infty}(1/b'_p)$, where $b'_{2p-1} = ib_{2p-1}$, $b'_{2p} = -ib_{2p}$, $p = 1, 2, 3, \cdots$. The sequence $\{b'_p\}$ evidently satisfies condition (H) if and only if the

sequence $\{b_p\}$ satisfies condition (H). We may therefore replace the conditions (3.3) by the conditions

$$(3.4) \quad k_1 > 0, k_{2p+1} \geq 0, I(k_{2p}) \leq 0, I(z_p) \geq \delta, |z_p| \leq M, p = 1, 2, 3, \dots$$

Proof of Theorem D. By Theorem A, it is necessary for the convergence of the continued fraction that the sequence $\{k_p\}$ satisfy condition (H). In case the series

$$(3.5) \quad \sum k_{2p+1}, \quad \sum k_{2p+1} |k_2 + k_4 + \dots + k_{2p}|^2$$

converge and

$$(3.6) \quad \lim_{p \rightarrow \infty} |k_2 + k_4 + \dots + k_{2p}| = \infty,$$

then the continued fraction converges by Theorem C. Hence there is to be considered here the case where at least one of the series (3.5) is divergent.

The inequality (3.2) is now

$$(3.7) \quad \left| \frac{A_p}{B_p} \right| \leq \frac{1}{k_1 \delta}, \quad p = 1, 2, 3, \dots,$$

so that $B_p \neq 0$, $p = 1, 2, 3, \dots$.

We have to consider two cases (A) and (B) according as $k_p \neq 0$ for all values of p , or $k_p = 0$ for one or more values of p .

(A) Suppose that $k_p \neq 0$ for $p = 1, 2, 3, \dots$. Put

$$(3.8) \quad c_0 = 0, \quad c_p = 1/k_p k_{p+1}, \quad p = 1, 2, 3, \dots,$$

$$(3.9) \quad T_0(w) = 1/k_1 w, \quad T_1(w) = c_1 + z_1 - \frac{c_1 c_2}{w},$$

$$T_p(w) = c_{2p-1} + c_{2p-2} + z_p - \frac{c_{2p} c_{2p-1}}{w}, \quad p = 2, 3, 4, \dots$$

One may then verify that the approximants of $\overset{\infty}{K}(1/b_p)$ are given by the formulas

$$(3.10) \quad \begin{aligned} T_0 T_1 \dots T_p (z_{p+1} + c_{2p} + c_{2p+1}) &= A_{2p+2}/B_{2p+2}, \\ T_0 T_1 \dots T_p (z_{p+1} + c_{2p}) &= A_{2p+1}/B_{2p+1}, \end{aligned} \quad p = 0, 1, 2, \dots$$

Under the hypothesis that (3.4) holds, we have

$$I(c_p) = \gamma_p \geq 0, \quad p = 0, 1, 2, \dots$$

Let

$$\beta_p = \gamma_{2p-2} + \gamma_{2p-1},$$

$$g_0 = 0, \quad g_p = \begin{cases} 1 & \text{if } \gamma_{2p+1} = 0, \\ \gamma_{2p}/(\gamma_{2p} + \gamma_{2p+1}) & \text{if } \gamma_{2p+1} > 0, \end{cases} \quad p = 1, 2, 3, \dots$$

$$\alpha_p = I[(c_{2p-1} c_{2p})^{1/2}],$$

Then we find that $0 \leq g_{p-1} \leq 1$ and $\alpha_p^2 = \gamma_{2p-1}\gamma_{2p} = \beta_p\beta_{p+1}(1 - g_{p-1})g_p$, $p = 1, 2, 3, \dots$. Thus the continued fraction generated by the sequence of transformations (3.9) is a *positive definite* continued fraction, (cf. [1] p. 265). Accordingly, we may construct a nest of circular regions $K_0 \supset K_1 \supset K_2 \supset K_3 \supset \dots$, where K_p is the image of the half-plane $I(w) \geq \beta_{p+1}g_p = \gamma_{2p} = I(1/k_{2p}k_{2p+1})$ under the transformation $t = T_0T_1 \dots T_p(w)$. Inasmuch as $I(z_{p+1} + c_{2p} + c_{2p+1}) > \gamma_{2p}$, $I(z_{p+1} + c_{2p}) > \gamma_{2p}$, we conclude by (3.10) that A_{2p+2}/B_{2p+2} and A_{2p+1}/B_{2p+1} have their values in K_p . To prove that the continued fraction $\prod_{p=1}^{\infty} (1/b_p)$ converges, it therefore suffices to prove that the radius r_p of K_p has the limit zero for $p = \infty$.

Let $u_0 = 1$, $u_p = (c_{2p-1}c_{2p})^{1/2}$, $v_p = c_{2p-2} + c_{2p-2p}$, $p = 1, 2, 3, \dots$;
 $Y_0(z) = 0$, $Y_1(z) = 1$,

$$-u_{p-1}Y_{p-1}(z) + (v_p + z_p)Y_p(z) - u_pY_{p+1}(z) = 0, \quad p = 1, 2, 3, \dots$$

Then the radius r_p of K_p satisfies the inequality

$$r_p \leq \frac{1}{2 \sum_{r=1}^p I(z_r) |Y_r(z)|^2} \leq \frac{1}{2\delta \sum_{r=1}^p |Y_r(z)|^2}.$$

This can be established by means of an easy extension of formula (3.12) of [1]. Hence $\lim_{p \rightarrow \infty} r_p = 0$ provided the infinite series $\sum_{r=1}^{\infty} |Y_r(z)|^2$ is divergent. Let $X_0(z) = -1$, $X_1(z) = 0$,

$$-u_{p-1}X_{p-1}(z) + (v_p + z_p)X_p(z) - u_pY_{p+1}(z) = 0, \quad p = 1, 2, 3, \dots$$

Then, as in formula (3.4) of [1], we have

$$\left| \frac{X_p(z)}{Y_p(z)} \right| \leq \frac{1}{I(z_1)} \leq \frac{1}{\delta}.$$

It follows that the series $\sum_{r=1}^{\infty} |Y_r(z)|^2$ diverges if the series $\sum_{r=1}^{\infty} |X_r(z)|^2$ diverges. Thus $\lim_{p \rightarrow \infty} r_p = 0$, provided at least one of the series

$$(3.11) \quad \sum_{r=1}^{\infty} |X_r(z)|^2, \quad \sum_{r=1}^{\infty} |Y_r(z)|^2$$

is divergent. Now one may prove by mathematical induction that when $z_p = 0$, $p = 1, 2, 3, \dots$, the series (3.11) are precisely the series (3.5). Hence the proof of Theorem D in the case under consideration may be immediately completed by means of the following *theorem of invariability*.

THEOREM 3.1. Let $u_0 = 1$, u_1, u_2, u_3, \dots be constants different from zero, and let v_1, v_2, v_3, \dots be constants. Let $x_p = X_p(z)$, $x_p = Y_p(z)$ be solutions of the equations $-u_{p-1}x_{p-1} + (v_p + z_p)x_p - u_px_{p+1} = 0$, $p = 1, 2, 3, \dots$ such that $X_0(z) = -1$, $X_1(z) = 0$, $Y_0(z) = 0$, $Y_1(z) = 1$. If the two series $\sum |X_p(z)|^2$ and $\sum |Y_p(z)|^2$ converge for $z_p = h_p$, $p = 1, 2, 3, \dots$, then these series converge uniformly for $|z_p - h_p| < M$, $p = 1, 2, 3, \dots$, where M is any positive constant.

This is a generalization of a theorem of Hellinger and Wall [3], and may be established by the method used in [3].

(B) Suppose that $k_p = 0$ for one or more values of p . Let

$$t_{2p}(w) = 1/(k_{2p} + w), \quad t_{2p-1}(w) = 1/(k_{2p-1}z_p + w), \quad p = 1, 2, 3, \dots$$

The n -th approximant of the continued fraction is given by $A_n/B_n = t_1 t_2 \dots t_n(0)$, $n = 1, 2, 3, \dots$.

We suppose first that there exists an index m such that

$$k_{2m-1} > 0, \quad k_{2m} + k_{2m+2} + \dots + k_{2m+2n} \neq 0 \text{ implies that } k_{2m+2n+1} = 0, \\ n = 0, 1, 2, \dots$$

Then, for all n sufficiently large,

$$\frac{A_n}{B_n} = t_1 t_2 \dots t_{2m-2} \left(\frac{1}{k_{2m-1}z_m + k_{2m+1}z_{m+1} + \dots + k_{2m+2r+1}z_{m+r+1} + w_n} \right),$$

where, if (3.4) holds, $I(w_n) \geq 0$, and where $r \rightarrow \infty$ as $n \rightarrow \infty$. Since

$$\sum_{p=1}^n k_{2p+1} |k_2 + k_4 + \dots + k_{2p}|^2 = \sum_{p=1}^{m-1} k_{2p+1} |k_2 + k_4 + \dots + k_{2p}|^2 \\ + |k_2 + k_4 + \dots + k_{2m-2}|^2 \sum_{p=m}^n k_{2p+1},$$

it follows that the first of the series (3.5) is divergent. We therefore conclude immediately that

$$\lim_{n \rightarrow \infty} \frac{A_n}{B_n} = \frac{A_{2m-2}}{B_{2m-2}},$$

which is finite by (3.7). The continued fraction is thus convergent in case such an index m exists.

If there is no index m with the specified property, then there is a continued fraction

$$(3.12) \quad \prod_{p=1}^{\infty} (1/b'_p), \quad b'_{2p-1} = k'_{2p-1}z'_p, \quad b'_{2p} = k'_{2p},$$

where k'_{2p-1} , k'_{2p} , z'_p are of the form

$$(3.13) \quad k'_{2p-1} = \sum_{r=\alpha}^{r=\beta} k_{2r+1} > 0, \quad k'_{2p} = \sum_{r=\beta}^{r=\gamma} k_{2r+2} \neq 0, \quad z'_p = \frac{\sum_{r=\alpha}^{r=\beta} k_{2r+1} z_{r+1}}{k'_{2p-1}},$$

whose sequence of approximants, $\{A'_n/B'_n\}$, is a subsequence of the sequence $\{A_n/B_n\}$. By (3.4) and (3.13), $I(k'_{2p}) \leq 0$, $I(z'_p) \geq \delta$. Also, at least one of the series $\sum k'_{2p+1}$ or $\sum k'_{2p+1} |k'_2 + k'_4 + \cdots + k'_{2p}|^2$ is divergent. Hence, by (A), the continued fraction (3.12) converges to a value L , the radii r'_p of the circles K'_p associated with (3.12) have the limit 0 for $p = \infty$, and their centers have the limit L for $p = \infty$. Those approximants in the sequence $\{A_n/B_n\}$ falling between A'_{2p}/B'_{2p} and A'_{2p+2}/B'_{2p+2} are all of the form $t'_1 t'_2 \cdots t'_{2p}(1/u)$, $t'_n(w) = 1/(b'_n + w)$, where $I(u) \geq 0$. Since $t'_1 t'_2 \cdots t'_{2p}(1/u)$ is the $2p$ -th approximant of (3.12) in which b'_{2p} has been replaced by $b'_{2p} + (1/u)$, it is given by the expression

$$(3.14) \quad T'_{p-1}(s), \quad s = z'_p + \frac{1}{k'_{2p-2}k'_{2p-1}} + \frac{1}{k'_{2p-1}(k'_{2p} + 1/u)};$$

where T'_{p-1} is the transformation defined in (3.9), formed for the continued fraction (3.12). Inasmuch as $I(s) \geq I(1/k'_{2p-2}k'_{2p-1})$, we conclude immediately that the value of (3.14) is in the circle K'_{p-1} . It therefore follows that $\lim_{n \rightarrow \infty} (A_n/B_n) = L$, so that the continued fraction is convergent.

This completes the proof of Theorem D.

Remark. The condition " $k_1 > 0$, $k_{2p+1} \geq 0$, $p = 1, 2, 3, \cdots$ " in Theorem D may be replaced by " $k_{2p-1} \geq 0$, $p = 1, 2, 3, \cdots$, with inequality for at least one p ." If, however, $k_{2p-1} = 0$, $p = 1, 2, 3, \cdots$, then the continued fraction is divergent.

As in the proof of Theorem C, we see that under the hypothesis of Theorem D, all the approximants of the continued fraction have their values in the circle

$$(3.15) \quad \left| w - \frac{1}{2k_1\delta} \right| = \frac{1}{2k_1\delta}.$$

Therefore, by Theorem D and a well known theorem on uniformly bounded sequences of analytic functions, we have the following theorem.

THEOREM E. If $k_1 > 0$, $k_{2p+1} \geq 0$, $R(k_{2p}) \geq 0$, $p = 1, 2, 3, \cdots$, and the sequence $\{k_p\}$ satisfies condition (H), then the continued fraction $\prod_{p=1}^{\infty} K(1/b_p)$, $b_{2p-1} = k_{2p-1}z$, $b_{2p} = k_{2p}$, converges uniformly over every bounded closed region in the right half-plane $R(z) > 0$, and its value is an analytic

function of z in this half-plane, which, for $R(z) \geq \delta > 0$, has its values in the circle (3.15).

4. Older theorems derived from Theorem D. We shall now obtain four older convergence theorems as easy consequences of Theorem D.

THEOREM 4.1. (Stieltjes [7]) *Let us suppose that $k_1 > 0$, $k_p \geq 0$, $p = 1, 2, 3, \dots$. If the series $\sum k_p$ diverges, then the continued fraction $\prod_{p=1}^{\infty} (1/b_p)$, $b_{2p-1} = k_{2p-1}z$, $b_{2p} = k_{2p}$, converges uniformly over every bounded closed region G of the z -plane whose distance from the negative half of the real axis is positive. If the series $\sum k_p$ converges, then the continued fraction diverges for every value of z .*

Proof. The sequence $\{k_p\}$ satisfies condition (H) if and only if the series $\sum k_p$ diverges, and therefore, by Theorem A, the divergence of this series is necessary for the convergence of the continued fraction. Let the distance of G from the negative half of the real axis be δ , and let $G = G_1 + G_2 + G_3$, where G_1 is in $R(z) \geq \delta$, G_2 is in $I(z) \geq \delta$, and G_3 is in $I(z) \leq -\delta$. If $\{k_p\}$ satisfies condition (H), the continued fraction converges uniformly over G_1 by Theorem E. Since

$$\prod_{p=1}^{\infty} (1/b_p) = \begin{cases} -i \cdot \prod_{p=1}^{\infty} K(1/b'_p), & b'_{2p-1} = k_{2p-1}(-iz), \quad b'_{2p} = ik_{2p}, \\ i \cdot \prod_{p=1}^{\infty} K(1/b''_p), & b''_{2p-1} = k_{2p-1}(iz), \quad b''_{2p} = -ik_{2p}, \end{cases}$$

the same conclusion holds for G_2 and G_3 , and therefore for G .

THEOREM 4.2. (Van Vleck [10]) *If $R(b_1) > 0$, $|I(b_p)| \leq cR(b_p)$, $p = 1, 2, 3, \dots$, where c is a positive constant, then the continued fraction $\prod_{p=1}^{\infty} K(1/b_p)$ converges if and only if the infinite series $\sum |b_p|$ is divergent.*

Proof. In Theorem D, take $k_{2p} = b_{2p}$, $k_{2p-1} = |b_{2p-1}|$, $k_{2p-1}z_p = b_{2p-1}$, where $z_p = 1$ if $b_{2p-1} = 0$. Then $R(k_{2p}) \geq |I(b_{2p})|/c \geq 0$, $k_1 = |b_1| > 0$, $k_{2p+1} = |b_{2p+1}| \geq 0$, $|z_p| = 1 < M$, $R(z_p) \geq 1/(1+c^2)^{1/2} = \delta$. If the series $\sum |b_p|$ diverges, then either $\sum k_{2p+1}$ diverges, or else $\sum |b_{2p}|$ diverges, in which case $\lim |k_2 + k_4 + \dots + k_{2p}| = \infty$ inasmuch as $|b_{2p}| \leq (1+c^2)^{1/2}R(b_{2p})$. We then conclude from Theorem D that the continued fraction converges if the series $\sum |b_p|$ diverges. If this series converges, then the continued fraction diverges by Theorem A.

Remark. Theorem 4.1 can be derived from Theorem 4.2, as Van Vleck showed.

If in Theorem 4.2 we require that $|I(b_p)| \leq cR(b_p)$ only for odd values of p , we obtain the following theorem.

THEOREM 4.3. (Mall [5]) *If $R(b_1) > 0$, $|I(b_{2p-1})| \leq cR(b_{2p-1})$, $R(b_{2p}) \geq 0$, $p = 1, 2, 3, \dots$, where c is a positive constant, then the continued fraction $\overset{\infty}{K}_{p=1} (1/b_p)$ converges if and only if the sequence $\{b_p\}$ satisfies condition (H).*

Proof. Choose the k_p and z_p as in the proof of Theorem 4.2, and apply Theorem D.

THEOREM 4.4. (Hamburger [2]) *If the k_p are real, $k_1 > 0$, $k_{2p+1} \geq 0$, $p = 1, 2, 3, \dots$, then the continued fraction of Theorem 4.1 converges for all nonreal z if and only if the sequence $\{k_p\}$ satisfies condition (H). When the condition is satisfied, the continued fraction converges uniformly over every closed bounded region whose distance from the real axis is positive.*

Proof. Use the transformations employed in the proof of Theorem 4.1, and apply Theorem E.

Remark. Theorems 4.1, 4.3 and 4.4 are somewhat more general than are to be found in the papers of Stieltjes, Mall and Hamburger, respectively.

The results found in this paper serve again to emphasize the importance of the theory of positive definite continued fractions. This theory overlaps or contains the greater part of the existing analytic theory of continued fractions. In this connection, we mention the following theorem of Thron [8].

THEOREM 4.5. *Let a and b be real numbers of which $a > 0$, and suppose that $I(b_p) \geq b$, $|I(a_p)| \leq a/2$, $p = 1, 2, 3, \dots$. If there exists a constant M such that $|a_p| < M$, $p = 1, 2, 3, \dots$, then the J -fraction $\overset{\infty}{K}_{p=1} (-a_{p-1}^2/(b_p + z))$, $a_0^2 = -1$, converges uniformly over every bounded closed region in the half-plane $I(z) > a - b$.*

As indicated (without proof) by Thron ([8], p. 413; cf. also [9], p. 779), this theorem is contained in the general theory of positive definite continued fractions. It is only necessary to put $z = (a - b)i + z'$, $\beta_p = I(b_p) + a - b$, $\alpha_p = I(a_p)$, and we see that $\beta_p \geq 0$, $\alpha_p^2 \leq \beta_p \beta_{p+1} \cdot \frac{1}{2}(1 - \frac{1}{2})$, $p = 1, 2, 3, \dots$,

so that the J -fraction of Thron is a particular positive definite J -fraction in the variable z' . It therefore converges for $I(z') > 0$, i. e. for $I(z) > a - b$, inasmuch as the series $\sum (1/|a_p|)$ diverges when $|a_p| < M$, $p = 1, 2, 3, \dots$

NORTHWESTERN UNIVERSITY

AND

THE UNIVERSITY OF TEXAS.

BIBLIOGRAPHY

1. J. J. Dennis and H. S. Wall, "The limit circle case for positive definite J -fractions," *Duke Mathematical Journal*, vol. 12 (1945), pp. 255-273.
2. H. Hamburger, "Ueber ein Erweiterung des Stieltjesschen Momentenproblems, Part II," *Mathematische Annalen*, vol. 82 (1920), pp. 120-164.
3. E. Hellinger and H. S. Wall, "Contributions to the analytic theory of continued fractions and infinite matrices," *Annals of Mathematics* (2), vol. 44 (1943), pp. 103-127.
4. H. von Koch, "Sur un théorème de Stieltjes et sur les fractions continues," *Bulletin de la Société Mathématique de France*, vol. 23 (1895), pp. 23-40.
5. J. Mall, "Ein Satz ueber die Konvergenz von Kettenbrüchen," *Mathematische Zeitschrift*, vol. 45 (1939), pp. 368-376.
6. M. A. Stern, *Lehrbuch der algebraischen Analysis*, Leipzig, 1860.
7. T. J. Stieltjes, *Recherches sur les fractions continues, Oeuvres*, vol. 2.
8. W. J. Thron, "Convergence regions for the general continued fraction," *Bulletin of the American Mathematical Society*, vol. 49 (1943), pp. 913-916.
9. ———, "A family of simple convergence regions for continued fractions," *Duke Mathematical Journal*, vol. 11 (1944), pp. 779-791.
10. E. B. Van Vleck, "On the convergence of continued fractions with complex elements," *Transactions of the American Mathematical Society*, vol. 2 (1901), pp. 215-233.

LEFT ASSOCIATED MATRICES WITH ELEMENTS IN AN ALGEBRAIC DOMAIN.*

By B. M. STEWART.

I. A Necessary and Sufficient Condition.

1. **Introduction.** For the set M of all n -by- n matrices, with elements in an algebraic field F of order k over the rational field Ra , all the usual definitions and properties of matrices with elements in an infinite field will hold. But for that subset $[M]$ of matrices whose elements are in the corresponding algebraic domain $[F]$ some of these properties have a special character. For example, every *non-singular* matrix P of M with determinant $d(P) \neq 0$ has an inverse matrix P^{-1} in M such that $PP^{-1} = P^{-1}P = I$, where I is the identity matrix of M ; but a matrix P of $[M]$, if it has an inverse matrix P^{-1} in $[M]$, must be not only non-singular, but also *unimodular*, i. e., the determinant $d(P)$ must be a unit of $[F]$.

As a further example, there is in the set M the notion of *rank*, but in the set $[M]$, the special notion of an *irreducible basis*. A matrix A of M is said to be of rank r if $n - r$ is the maximum number of zero rows to be found in any matrix PA where P is non-singular in M ; a matrix A of $[M]$ is said to have an irreducible basis of b rows if $n - b$ is the maximum number of zero rows to be found in any matrix PA where P is unimodular in $[M]$. As will be described in §7, if the domain $[F]$ is a principal ideal ring, then b is always equal to r ; but if $[F]$ is not a principal ideal ring, then there are matrices for which $b = r + 1$. This anomaly is of importance in the present problem.

Two matrices A and B of $[M]'$ are said to be left associates in $[M]$ if there exists a unimodular matrix P of $[M]$ such that $PA = B$. This notion is an equals relation dividing the matrices of $[M]$ into mutually exclusive classes of left associated matrices. Our fundamental problem is to determine necessary and sufficient conditions for this class division.

If the domain under consideration is a principal ideal ring, a necessary and sufficient condition that A and B be left associates is that A and B have

* Received August 7, 1946; Presented to the Society, in part, April 12, 1941, as a portion of a thesis written at the University of Wisconsin. The partial results announced earlier are herein completely extended.

the same Hermite triangular canonical form (zeros above the main diagonal; if a diagonal element is zero, then its row is zero; elements below the diagonal in a prescribed set of residues modulo the diagonal element above).¹ However, this test is a practical one only if there is an algorithm for determining a greatest common divisor, such as the Euclidean algorithm for the rational domain $[Ra]$; and for domains whose class number is greater than one, the presence of non-principal ideals prevents the *direct* solution of the problem by an Hermite form.

However, if each element of a matrix A of $[M]$ is replaced by its k -by- k second matrix representation, there is produced an *enlarged* matrix A^E , of order kn -by- kn , to be sure, but with elements in the rational domain! Hence for A^E the Hermite form is well-defined *and* easily found. In this paper we prove that a necessary and sufficient condition that matrices A and B of $[M]$ be left associates in $[M]$ is that the corresponding enlarged matrices A^E and B^E be left associates, i. e., that the enlarged matrices have the same Hermite form. Although the proof and the actual construction of a unimodular transforming matrix P (whose construction we defer to Part II) depend upon other considerations, the test itself involves only the rational domain and affords a practical criterion for the division of the matrices of $[M]$ into classes of left associated matrices.

The proof is accomplished by considering the more primitive concept of *mutual left divisibility*: two matrices A and B of $[M]$ are called mutually left divisible if there exist matrices P and Q of $[M]$ such that $PA = B$ and $QB = A$. We adapt from the literature the proof that mutual left divisibility of A and B in $[M]$ is equivalent to left associativity of A and B in $[M]$. We prove that mutual left divisibility of A and B in $[M]$ is equivalent to the left associativity of A^E and B^E . Hence our principal result follows.

Our problem is related to the papers of Steinitz,² Krull,³ and Franz⁴ who have solved the problems of finding for matrices of $[M]$ the necessary and sufficient conditions for *divisibility*: $PAQ = B$; for *mutual divisibility*: $PAQ = B$ and $SBT = A$; and for *equivalence*: $PAQ = B$, with P and Q

¹ C. C. MacDuffee, "Matrices with elements in a principal ideal ring," *Bulletin of the American Mathematical Society*, vol. 39 (1933), pp. 564-584.

² E. Steinitz, Recteckige Systeme und Moduln in algebraischen Zahlkörpern," *Mathematische Annalen*, vol. 71 (1911), pp. 328-354, and vol. 72 (1912), pp. 297-345.

³ W. Krull, "Matrizen, Moduln und verallgemeinerte Abelsche Gruppen in Bereich der ganzen algebraischen Zahlen," *Sitzungsberichte Heidelberg Akademie der Wissenschaften*, 2 Abhandlung (1932), pp. 13-18.

⁴ W. Franz, "Elementarteilertheorie in algebraischen Zahlkörpern," *Journal für die reine und angewandte Mathematik*, vol. 171 (1934), pp. 149-161.

both unimodular. Obviously left associativity is a special type of equivalence. With these authors, we have no hesitation in using a canonical form enlarged beyond that used when the domain is a principal ideal ring. The new feature is that our canonical form has elements not in the original domain $[F]$, but in the rational domain $[Ra]$.

2. Second matrix representation. If the algebraic field F is of order k , then a set of k elements, $e_1 = 1, e_2, \dots, e_k$, in the domain $[F]$, can be found to serve as a *basis* for the field in the sense that for every element a of F there is a unique set of k numbers a_i in the rational field Ra such that

$$(1) \quad a = a_1 e_1 + a_2 e_2 + \dots + a_k e_k;$$

further, a is in $[F]$, if and only if each a_i of (1) is in the rational domain $[Ra]$.⁵

Each product $e_i e_j$ is in $[F]$, and hence can be written in the form

$$(2) \quad e_i e_j = \sum_{m=1}^k c_{ijm} e_m, \quad (i, j = 1, \dots, k),$$

where the k^3 constants of multiplication c_{ijm} are in $[Ra]$.

If we define S_i to be the k -by- k matrix $S_i = (c_{ris})$, then the correspondence

$$a \leftrightarrow S(a) = a_1 S_1 + a_2 S_2 + \dots + a_k S_k,$$

where the a_i are the same as in (1), defines the so-called *second matrix representation* of F , for it is well-known that the system of k -by- k matrices $S(a)$ with matrix addition and multiplication as operations is isomorphic with the field F .

Since $e_1 = 1$, the first row of S_i has $c_{1is} = d_{is}$ where $d_{is} = 0$ if $i \neq s$ and $d_{ii} = 1$. Thus the first row of $S(a)$ is exactly (a_1, a_2, \dots, a_k) . In other words, the matrix $S(a)$ is completely determined by its first row. In particular, $S(a)$ represents a number of $[F]$ if and only if all the elements of $S(a)$ are in $[Ra]$. These facts provide the basis for our fundamental lemma.

LEMMA. *If P is a k -by- kn matrix with elements in $[Ra]$ such that*

⁵ For example, see E. Landau, *Einführung in die elementare und analytische Theorie der algebraischen Zahlen und der Ideale*, Part I, Teubner (1918).

$$P \begin{pmatrix} S(a_1) \\ \cdot \\ \cdot \\ \cdot \\ S(a_n) \end{pmatrix} = S(b),$$

where a_1, a_2, \dots, a_n, b are in $[F]$, then there exist n numbers p_1, p_2, \dots, p_n in $[F]$ such that

$$(S(p_1) \cdots S(p_n)) \begin{pmatrix} S(a_1) \\ \cdot \\ \cdot \\ \cdot \\ S(a_n) \end{pmatrix} = S(b).$$

Separate the first row of P into n blocks each 1-by- k , say $p_{11}, \dots, p_{1k}; p_{21}, \dots, p_{2k}; \dots; p_{n1}, \dots, p_{nk}$; and define $p_i = p_{i1}e_1 + \dots + p_{ik}e_k$, $i = 1, 2, \dots, n$. Evidently p_i is in $[F]$, for each p_{ij} is in $[Ra]$. The first row of the matrix $(S(p_1) \cdots S(p_n))$ is the same as the first row of the matrix P ; hence, by the hypothesis concerning the given matrix product, the new matrix product agrees in its first row with the first row of $S(b)$. But the new product considered in terms of k -by- k blocks is the sum and product of S -matrices and is therefore, by the isomorphism explained above, an S -matrix. Finally, since an S -matrix is completely determined by its first row, the new product must be exactly $S(b)$.

3. Enlarged matrices. If each element of a matrix A of M is replaced by its second matrix representation there is produced an *enlarged matrix* A^E which is kn -by- kn with elements in Ra . Conversely, to a matrix A^E which is kn -by- kn with elements in Ra and each of whose k -by- k blocks is in S -form there corresponds an n -by- n matrix A of M .

If A, B, C are matrices of M such that $AB = C$, then $A^E B^E = C^E$; and conversely. This follows from the possibility of block-multiplication of matrices.

If A is non-singular in M , then A^E is non-singular. For $A.A^{-1} = I$ implies $A^E(A^{-1})^E = I^E$, hence $(A^E)^{-1} = (A^{-1})^E$.

If A is of rank r in M , then A^E is of rank kr . For there exist non-singular matrices P and Q of M such that

$$PAQ = \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}. \quad \text{Then} \quad P^E A^E Q^E = \begin{pmatrix} I_r^E & 0^E \\ 0^E & 0^E \end{pmatrix}.$$

But P^E and Q^E are non-singular, hence rank-preserving. Therefore the rank of A^E is the same as the rank of I_r^E .

If A is in $[M]$, then all the elements of A^E are in $[Ra]$; and conversely. For if $A = (a_{rs})$, then all the elements of $S(a_{rs})$ are in $[Ra]$ if and only if a_{rs} is in $[F]$.

If P is unimodular in $[M]$, then P^E is unimodular in $[Ra]$; and conversely. For if P^{-1} is in $[M]$, then $(P^{-1})^E = (P^E)^{-1}$ has all elements in $[Ra]$; and conversely.

4. Steinitz: On mutual left divisibility. From the work of Steinitz we borrow the following theorem:

If A, B, U, V are matrices of $[M]$ such that $UA = B$ and $VB = A$, then there exists a unimodular matrix P in $[M]$ such that $PA = B$.

The converse is evident. Therefore, using the terms of 1, we may state that the matrices A and B of $[M]$ are left associates in $[M]$ if and only if they are mutually left divisible in $[M]$.

5. Left associated enlarged matrices. We are now in a position to establish the following theorem:

A necessary and sufficient condition that the matrices A and B of $[M]$ be mutually left divisible in $[M]$ is that the enlarged matrices A^E and B^E be left associates in $[Ra]$.

If A and B are mutually left divisible in $[M]$, then there exist matrices U and V of $[M]$ such that $UA = B$ and $VB = A$. Then by 4, there exists a unimodular matrix P of $[M]$ such that $PA = B$. Hence by 3 we have $P^E A^E = B^E$ with P^E unimodular in $[Ra]$. Therefore A^E and B^E are left associates in $[Ra]$.

Conversely, if A^E and B^E are left associates in $[Ra]$, then they have the same Hermite form H and there exist matrices X and Y , kn -by- kn and unimodular in $[Ra]$, such that $XA^E = H = YB^E$. Then $U' = Y^{-1}X$ and $V' = X^{-1}Y$ are unimodular in $[Ra]$ and such that $U'A^E = B^E$ and $V'B^E = A^E$. However, the matrices U' and V' do not, in general, have their k -by- k blocks in S -form. Therefore by repeated applications of the fundamental lemma in 2, we construct new transforming matrices that will be in S -form. We separate the matrices U' and V' into k -by- kn row blocks. To each of these we apply the lemma, i. e., using only the *first* rows of these row blocks, we define the rows of two matrices of $[M]$ which we shall designate as U and V .

By the lemma, these new matrices are such that $U^E A^E = B^E$ and $V^E B^E = A^E$. Therefore U and V are such that $UA = B$ and $VB = A$. Hence A and B are mutually left divisible in $[M]$.

Although U' and V' are unimodular in $[Ra]$, it does not follow, in general, that the matrices U^E and V^E , derived from U' and V' in the manner of the lemma, enjoy the same property. Therefore we cannot at once conclude that A and B are left associates. But by combining this theorem with the theorem of 4, we have proved the following:

THEOREM I. *A necessary and sufficient condition that the matrices A and B of $[M]$ be left associates in $[M]$ is that the enlarged matrices A^E and B^E have the same Hermite form.*

Having applied this test to a pair of matrices A and B of $[M]$ and having discovered that A^E and B^E have the same Hermite form H , so that the existence of a matrix P , unimodular in $[M]$ and such that $PA = B$, is known, we may demand an explicit construction of the transforming matrix P . Toward this end we must distinguish various cases according to the rank of A and B . The rank of A and B can be immediately determined from the rank of H , for as we have noted in 3 the rank of the enlarged matrices is k -fold that of the original matrices. More than this can be said, for in the process of finding H we work on the columns of A^E from right to left; if we pause after each set of k columns, we must find that either k or no independent columns have been added to H , as will be evidenced by a set of k non-zero or a set of k zero entries, respectively, in the principal diagonal of H . By reference to H we can determine the independent columns of A and B , considered from right to left and occurring in the same positions for both A and B . Then, in particular, we may suppose, for subsequent convenience of notation, that the corresponding columns of A and B are rearranged so that, if the common rank is r , there are r independent columns of A and B at the extreme right of the matrices. Matrices A and B , thus arranged, will be said to be in *preferred form*. A matrix A can be recognized to be in preferred form if the Hermite form H of A^E has all the non-zero elements of the principal diagonal in the kr places at the extreme right, so that

$$(3) \quad H = \begin{pmatrix} 0 & 0 \\ H_2 & H_1 \end{pmatrix}$$

where H_1 is kr -by- kr , triangular, and of rank kr .

II. Construction of the Transforming Matrix.

6. The non-singular case. If the matrices A and B of $[M]$ are non-singular, we can give a direct proof, independent of the Steinitz theory, that the matrices A and B are left associates in $[M]$ if and only if the enlarged matrices have the same Hermite form. At the same time we find an explicit construction within $[Ra]$ for the transforming unimodular matrix P of $[M]$ such that $PA = B$.

Suppose A^E and B^E have the same Hermite form H so that there exist matrices X and Y , kn -by- kn , unimodular in $[Ra]$, such that $XA^E = H = YB^E$. Then $P' = Y^{-1}X$ is unimodular in $[Ra]$ and $P'A^E = B^E$. Since A is non-singular in M , A^E is non-singular (see 3) and $(A^E)^{-1} = (A^{-1})^E$. Therefore $P' = B^E(A^E)^{-1} = B^E(A^{-1})^E$, and we see that P' has its k -by- k blocks in S -form and hence can be written $P' = P^E$. Since P' has all its elements in $[Ra]$, P is in $[M]$. Furthermore, P' is unimodular in $[Ra]$, hence P is unimodular in $[M]$ (see 3). From $P'A^E = P^EA^E = B^E$ it follows that $PA = B$. Therefore A and B are left associates in $[M]$.

Conversely, if A and B are left associates in $[M]$ so that $PA = B$ with P unimodular in $[M]$, then $P^EA^E = B^E$ with P^E unimodular in $[Ra]$, so A^E and B^E are left associates in $[Ra]$ and must have the same Hermite form.

THEOREM II. *If A and B are non-singular matrices of $[M]$ such that $XA^E = H = YB^E$ has rank kn , then the matrix P , unimodular in $[M]$, such that $PA = B$, is given by both $P^E = Y^{-1}X$ and $P^E = B^E(A^E)^{-1}$.*

7. The singular case with principal column class. If the matrices A and B of $[M]$ are singular, the fact that A^E and B^E have the same Hermite form guarantees the existence of a matrix P , unimodular in $[M]$, such that $PA = B$, but a direct construction of the transforming matrix P seems to be attended with a number of difficulties, as forewarned in 1.

Using the notion of the division of the ideals of $[F]$ into classes of equivalent ideals, we can, after a few preliminary definitions, state Steinitz' elegant theorem, exactly descriptive of the situation.

Let $A^{(m)}$ indicate the m -th *adjugate* of A , i. e., the matrix of order $\binom{n}{m}$ formed by arranging all the determinants of order m that can be formed from A in lexicographic order as to rows and columns. If A is of rank r , then $A^{(r)}$ is of rank 1; hence the columns of $A^{(r)}$ are proportional; hence

the ideals determined by the different columns of $A^{(r)}$ belong to the same class of ideals, described as the *column class* of A .

If A is in $[M]$ and if $n - b$ is the maximum number of zero rows to be found in any matrix QA , where Q is unimodular in $[M]$, then the matrix A is said to possess an *irreducible basis* of b rows.

The *principal class* is the class of ideals of $[F]$ consisting of all the principal ideals of $[F]$.

In terms of these definitions the theorem of Steinitz is as follows:

If the matrix A of $[M]$ is of rank r , then A has an irreducible basis of $b = r$ or $b = r + 1$ rows, according as the column class of A is or is not the principal class.

Only an outline of the proof will be presented here.⁶ A row q of n elements of $[F]$ is said to belong to the left null-space of A if $qA = 0$. If the rank r of A is such that $r \leq n - 2$ so that the left null-space of A has at least two independent rows, then it can be proved that the left null-space of A contains a row of relatively prime numbers of $[F]$; it can be proved that this latter row can be augmented to a matrix Q unimodular in $[M]$ and such that QA has at least one zero row. Continuing in this manner, we have the result that in all cases the irreducible basis has either $b = r$ or $b = r + 1$ rows, the proof being obvious in the omitted cases $r = n$ and $r = n - 1$. The proof is concluded by identifying the case $b = r$ with a matrix A whose column class is the principal class.

In some of these operations the method of enlarged matrices is helpful. For example, in considering the left null-space of A , we can use $XA^E = H$; for from any one of the $k(n - r)$ rows of X with elements in $[Ra]$ which produce a zero row of H , using the lemma of 2, we can immediately derive a row with elements in $[F]$ belonging to the left null-space of A . It does not follow, however, that two distinct rows of X which are, of course, independent in Ra , will have derived rows independent in F .

We can also give a method based on enlarged matrices for deciding whether the elements of a row are relatively prime; for we can give an explicit construction for the canonical minimal basis of any ideal of $[F]$. Consider, for example, the ideal (a_1, \dots, a_n) . If the Hermite form of the kn -by- k column vector formed from $S(a_1), \dots, S(a_n)$ has in its last k -by- k block the elements h_{ij} , with $h_{ij} = 0$ if $j > i$, then $h_1 = h_{11}e_1$, $h_2 = h_{21}e_1 + h_{22}e_2$, \dots , $h_k = h_{k1}e_1 + \dots + h_{kk}e_k$ form the unique canonical

⁶ For example, see Franz, *loc. cit.*, Hilfsatz 3, Satz 5, Satz 6.

minimal basis for the ideal.⁷ In particular, if $(a_1, \dots, a_n) = (1)$, then the last k -by- k block of the Hermite form must be $S(1) = I_k$. Furthermore, there is a direct construction for the numbers p_1, \dots, p_n of $[F]$ such that $p_1 a_1 + \dots + p_n a_n = 1$. For if P is the matrix unimodular in $[Ra]$ such that

$$P \begin{pmatrix} S(a_1) \\ \cdot \\ \cdot \\ \cdot \\ S(a_n) \end{pmatrix} = \begin{pmatrix} 0 \\ \cdot \\ \cdot \\ 0 \\ I \end{pmatrix}$$

then from the $k(n-1) + 1$ -st row of P we may derive, in the manner of the lemma of 2, the values of p_1, \dots, p_n .

Unfortunately the Steinitz process of augmenting a row of relatively prime numbers of $[F]$ to a unimodular matrix of $[M]$ is more complicated for the general algebraic domain $[F]$ than it is for the well-known case of the rational domain $[Ra]$, and it does not seem that the method of enlarged matrices is particularly helpful in this fundamental step.

As a corollary to this Steinitz theorem we observe that if the domain $[F]$ is a principal ideal ring, then the column class of each matrix A must be the principal class and hence for each A we find $b = r$, the familiar situation in $[Ra]$.

By the method of enlarged matrices we have noted how to construct for any ideal a minimal basis of k elements, i. e., a basis with respect to $[Ra]$. As a further corollary to the Steinitz theorem we note that an ideal can be written as a column vector of rank 1, and therefore has an irreducible basis, i. e., a basis with respect to $[F]$, of one or two elements, according as the ideal is principal or non-principal.

In conclusion to this section we have the following result:

THEOREM III. *If A and B are singular matrices of $[M]$ of rank r whose enlarged matrices have the same Hermite form, whose column class is the principal class, and which are in preferred form, then the matrix P unimodular in $[M]$ such that $PA = B$ is given by $P = R^{-1}P'Q$. The matrices Q and R are unimodular in $[M]$ such that the last r rows of $QA = A'$ and $RB = B'$ are the only non-zero rows. The matrix P' unimodular in $[M]$ is determined by $(P')^E = Y^{-1}X$ where $X(A')^E = H = Y(B')^E$ with matrices as follows:*

⁷ MacDuffee, *loc. cit.* in 1.

$$(A')^E = \begin{pmatrix} 0 & 0 \\ A^{E_2} & A^{E_1} \end{pmatrix}, (B')^E = \begin{pmatrix} 0 & 0 \\ B^{E_2} & B^{E_1} \end{pmatrix}, H = \begin{pmatrix} 0 & 0 \\ H_2 & H_1 \end{pmatrix}$$

$$X = \begin{pmatrix} I & 0 \\ 0 & X_1 \end{pmatrix}, Y = \begin{pmatrix} I & 0 \\ 0 & Y_1 \end{pmatrix}, (P')^E = \begin{pmatrix} I & 0 \\ 0 & P^{E_1} \end{pmatrix},$$

where P^{E_1} is given by both $P^{E_1} = Y^{-1}_1 X_1$ and $P^{E_1} = B^{E_1} (A^{E_1})^{-1}$.

8. The singular case with a non-principal column class. If the matrices A and B of $[M]$ are singular of rank r with a non-principal column class, then there exist unimodular matrices Q and R of $[M]$ such that the last $r+1$ rows of $QA = A'$ and $RB = B'$ are the only non-zero rows. If $X'(A')^E = H = Y'(B')^E$, then we succeed, as described in 5, in constructing matrices U and V of $[M]$ such that $UA' = B'$ and $VB' = A'$. The theorem of Steinitz, given in 4, then guarantees the existence of a matrix P' , unimodular in $[M]$, such that $P'A' = B'$. In the cases of matrices that are non-singular or are singular with a principal column class the discussions of 6 and 7 provide a construction for P' . But in the case here under discussion a very different attack seems to be required; we shall outline Steinitz' proof and indicate a number of steps in which the use of enlarged matrices is helpful. Having found P' , we can construct $P = R^{-1}P'Q$ unimodular in $[M]$ and such that $PA = B$.

If A and B are in preferred form, then

$$A' = \begin{pmatrix} 0 & 0 \\ A_2 & A_1 \end{pmatrix}, \quad B' = \begin{pmatrix} 0 & 0 \\ B_2 & B_1 \end{pmatrix},$$

where A_1 and B_1 are $r+1$ -by- $r+1$ and of rank r . Therefore

$$X' = \begin{pmatrix} I & 0 \\ 0 & X_1 \end{pmatrix}, \quad Y' = \begin{pmatrix} I & 0 \\ 0 & Y_1 \end{pmatrix},$$

where X_1 and Y_1 are $k(r+1)$ -by- $k(r+1)$ and unimodular in $[Ra]$. Matrices U_1 and V_1 can therefore be derived from $Y^{-1}_1 X_1$ and $X^{-1}_1 Y_1$, respectively, as explained in 5, so that

$$U = \begin{pmatrix} I & 0 \\ 0 & U_1 \end{pmatrix}, \quad V = \begin{pmatrix} I & 0 \\ 0 & V_1 \end{pmatrix}.$$

The essential problem reduces to being given matrices A_1 and B_1 of order $r+1$ and of rank r with elements in $[F]$ and mutually left divisible, since $U_1 A_1 = B_1$ and $V_1 B_1 = A_1$, and being asked to prove that A_1 and B_1 are left associated, i. e., that there exists a matrix P_1 of order $r+1$ and unimodular in $[F]$ such that $P_1 A_1 = B_1$. For having found P_1 we may take

$$P' = \begin{pmatrix} I & 0 \\ 0 & P_1 \end{pmatrix}.$$

Since A_1 is of order $r+1$ and rank r , the left null-space of A_1 has a basis of one row $w_1 = w_{11}, \dots, w_{1,r+1}$. Since the first k rows of $X_1 A_1^E = H_1$ are zero rows, it is easy to derive, as in the lemma of 2, a row with elements in $[F]$ to serve as w_1 . Let T_1 indicate the ideal determined by the elements of w_1 . Any other row $w_2 = w_{21}, \dots, w_{2,r+1}$ in the left null-space of A_1 and with elements in $[F]$ determines an ideal T_2 in the same ideal class as T_1 because the row w_2 is proportional to the row w_1 . Since in every ideal class there exist ideals relatively prime to any given ideal,⁸ the row w_2 can be determined with elements in $[F]$ so that $(T_1, T_2) = (1)$. It is then possible to express any other row $w_3 = w_{31}, \dots, w_{3,r+1}$ of the left null-space of A_1 with elements in $[F]$ in the form $w_3 = m_1 w_1 + m_2 w_2$ where m_1 and m_2 are in $[F]$. For in terms of ideal factors we have $w_{1i} = T_1 W_i$, $w_{2i} = T_2 W_i$, $(w_{1i}, w_{2i}) = (T_1 W_i, T_2 W_i) = (T_1, T_2) W_i = W_i$; hence $w_{3i} = T_3 W_i$ which is in the ideal W_i can be written $w_{3i} = m_1 w_{1i} + m_2 w_{2i}$ where m_1 and m_2 are in $[F]$. The same coefficients m_1 and m_2 serve for $i = 1, \dots, r+1$ because the rows w_1, w_2, w_3 are proportional.

Consider any row w with elements in $[F]$, but not necessarily in the left null-space of A_1 . From $wA_1 = wV_1B_1 = wV_1U_1A_1$ it follows that the row $w - wV_1U_1$ is in the left null-space of A_1 and can be written $w - wV_1U_1 = m_1 w_1 + m_2 w_2$. Hence w is a linear combination with coefficients in $[F]$ of the rows w_1 and w_2 and the rows of U_1 . If this result is applied, successively, to the rows of the identity matrix I of order $r+1$, we see that there exists a matrix Z which is $r+1$ -by- $r+3$ with elements in $[F]$ such that

$$I = Z \begin{pmatrix} U_1 \\ w_1 \\ w_2 \end{pmatrix}.$$

In terms of the $r+1$ -st adjugates we find

$$1 = I^{(r+1)} = Z^{(r+1)} \begin{pmatrix} U_1 \\ w_1 \\ w_2 \end{pmatrix}^{(r+1)};$$

so that (1) is the greatest common ideal divisor of the determinants of the $r+1$ -st order that can be formed from $\begin{pmatrix} U_1 \\ w_1 \\ w_2 \end{pmatrix}$. Since w_1 and w_2 are pro-

⁸ Steinitz, *loc. cit.*, Part I, 1.

portional many of these $r+1$ -st order determinants are zero, and the result may be written

$$(4) \quad (1) = (d(U_1), d(M_1), \dots, d(M_{r+1}), d(N_1), \dots, d(N_{r+1})),$$

where M_i is the matrix formed by replacing the i -th row of U_1 by w_1 , and N_i is the matrix formed by replacing the i -th row of U_1 by w_2 . If $U_1 = (u_{ij})$, let U_{ij} denote the cofactor of u_{ij} , so that $d(M_i) = \sum_j w_{1j} U_{ij}$ and $d(N_i) = \sum_j w_{2j} U_{ij}$. Using these formulas and well-known properties of the cofactors, we can easily establish the following results:

$$(5) \quad \begin{aligned} \sum_i d(M_i) u_{it} &= w_{1t} d(U_1), \\ \sum_i d(N_i) u_{it} &= w_{2t} d(U_1), \end{aligned} \quad t = 1, \dots, r+1.$$

Suppose that $d(M_1), \dots, d(M_{r+1}), d(N_1), \dots, d(N_{r+1})$ have a common ideal divisor T . From (4) it follows that either $T = (1)$ or that T does not divide $d(U_1)$. In the latter case it follows from (5) that T divides w_{1i} and w_{2i} for $i = 1, \dots, r+1$. But $(w_{11}, \dots, w_{1,r+1}, w_{21}, \dots, w_{2,r+1}) = (T_1, T_2) = (1)$; hence again $T = (1)$. Therefore

$$(1) = (d(M_1), \dots, d(M_{r+1}), d(N_1), \dots, d(N_{r+1})).$$

Using the method of enlarged matrices, described in 7, for determining a minimal basis of an ideal, we have an explicit method for determining the numbers $a'_1, \dots, a'_{r+1}, b'_1, \dots, b'_{r+1}$ of $[F]$ such that

$$(6) \quad a'_1 d(M_1) + \dots + a'_{r+1} d(M_{r+1}) + b'_1 d(N_1) + \dots + b'_{r+1} d(N_{r+1}) = 1.$$

Let us define $a_i = (1 - d(U_1))a'_i$, $b_i = (1 - d(U_1))b'_i$, $i = 1, \dots, r+1$; and

$$(7) \quad P_1 = U_1 + \begin{bmatrix} a_1 w_1 \\ \vdots \\ a_{r+1} w_1 \end{bmatrix} + \begin{bmatrix} b_1 w_2 \\ \vdots \\ b_{r+1} w_2 \end{bmatrix}.$$

Well-known properties of determinants enable us to expand

$$d(P_1) = d(U_1) + \sum_i a_i d(M_i) + \sum_i b_i d(N_i);$$

by virtue of (6) we conclude that $d(P_1) = 1$; hence P_1 is a matrix unimodular in $[F]$. Furthermore, since $w_1 A_1 = 0 = w_2 A_1$, we find $P_1 A_1 = U_1 A_1 = B_1$. Hence A_1 and B_1 are left associates as was to be proved.

Those steps of the above proof which would be essential in the construction of the transforming matrix can be summarized, as follows:

THEOREM IV. *If A and B are singular matrices of $[M]$ of rank r whose enlarged matrices have the same Hermite form, whose column class is not the principal class, and which are in preferred form, then the matrix P unimodular in $[M]$ such that $PA = B$ is given by $P = R^{-1}P'Q$. The matrices Q and R are unimodular in $[M]$ such that the last $r + 1$ rows of $QA = A'$ and $RB = B'$ are the only non-zero rows. From $X'(A')^E = H = Y'(B')^E$ determine X_1 and Y_1 . From $Y_1^{-1}X_1$ derive U_1 such that $U_1A_1 = B_1$. From one of the first k rows of X_1 derive a row w_1 such that $w_1A_1 = 0$. Determine a row w_2 such that $w_2A_1 = 0$ and $(T_1, T_2) = (1)$, where $T_i = (w_{i1}, \dots, w_{i, r+1})$. Compute $d(M_1), \dots, d(N_{r+1})$, where M_i and N_i are the matrices obtained from U_1 by replacing the i -th row by w_1 and w_2 , respectively. Using enlarged matrices, from*

$$N \begin{bmatrix} S(d(M_1)) \\ \vdots \\ S(d(N_{r+1})) \end{bmatrix} = S(1)$$

derive the numbers a'_1, \dots, b'_{r+1} satisfying (6). Construct P_1 as in (7) and take

$$P' = \begin{pmatrix} I & 0 \\ 0 & P_1 \end{pmatrix}.$$

MICHIGAN STATE COLLEGE,
EAST LANSING, MICHIGAN.

EXTENSIONS OF HARMONIC TRANSFORMATIONS.*

By EDWARD KASNER and JOHN DE CICCIO.

1. **Harmonic transformations.** We shall study the transformation theory of differential elements of third order, all of which pass through a fixed point, which is induced by the infinite set of harmonic transformations.¹ Also we shall consider the transformation theory of third order differential elements under the group of arbitrary point transformations. From this work, a comparison of the two theories may be deduced.

We shall term a correspondence T of the (x, y) -plane a *harmonic transformation* if it is defined by the equations

$$(1) \quad X = \phi(x, y), \quad Y = \psi(x, y),$$

where the jacobian

$$(2) \quad J = \phi_x \psi_y - \phi_y \psi_x \neq 0,$$

such that the components ϕ and ψ obey the Laplace equation

$$(3) \quad \phi_{xx} + \phi_{yy} = 0, \quad \psi_{xx} + \psi_{yy} = 0,$$

but otherwise are not interrelated in any way whatsoever.

Harmonic transformations should not be confused with conformal correspondences. Conformal maps are those for which the components ϕ and ψ obey the direct or reverse Cauchy-Riemann equations

$$(4) \quad \phi_x = \pm \psi_y, \quad \phi_y = \mp \psi_x.$$

The components ϕ and ψ of a harmonic transformation do not satisfy any differential conditions of first order in general, and therefore a harmonic transformation is not usually conformal. Thus the conformal group is a proper subset of the infinite set of harmonic transformations.

The group of arbitrary point transformations induces the general projective group of three parameters between the pencils of lineal elements whose vertices are at the points corresponding under a transformation T . *The infinite set (II) of harmonic transformations induces the same three-*

* Received December 28, 1946; Presented to the American Mathematical Society, 1946.

¹ Kasner and De Ciccio, "Theory of harmonic transformations," *Proceedings of the National Academy of Sciences*, vol. 33 (1947), pp. 20-23.

parameter projective group. The conformal group induces a one-parameter group, namely, the group of rotations.

By the group of arbitrary point transformations, there corresponds an eight-parameter group G_8 between the bundles of differential elements of second order whose vertices are at the points corresponding under the transformation T' . Kasner has presented a complete study of this group,² obtaining all the possible invariants and also classifying all the differential equations of second order which are algebraic in the derivatives by means of this group G_8 . *By the infinite set (H) of harmonic transformations, there corresponds the same eight-parameter group G_8 between the bundles of second order differential elements.* By the conformal group, there corresponds a four-parameter group G_4 .

The group of arbitrary point transformations induces on the differential element of third order a fifteen-parameter group G_{15} . This group G_{15} is equivalent to projective geometry on a three-dimensional manifold M_5 of fifth degree embedded in projective space R_{11} of eleven dimensions. By means of this fifteen-parameter group G_{15} , we obtain a classification of differential equations of third order which are algebraic in the derivatives.

The infinite set of harmonic transformations induces on the differential elements of third order a twelve-parameter set S_{12} of transformations. This set S_{12} is not a group. Thus the first essential difference in the extensions of the group of arbitrary point transformations and those of the infinite set of harmonic transformations is in the third order of differentiation.

It is noted that by the conformal group, there corresponds a six-parameter group G_6 of transformations of third order differential elements, which we have studied elsewhere.³

2. Transformation theory of differential elements of second order.

In order to simplify our work, we shall use the minimal coordinates $u = x + iy$

² Kasner, "The geometry of differential elements of the second order with respect to the group of all point transformations," *American Journal of Mathematics*, vol. 28 (1906), pp. 203-213.

³ Kasner and De Ciccio, "Conformal geometry of third order differential elements," *Revista de la Universidad Nacional de Tucuman (Argentina)*, vol. 2 (1941), pp. 51-58. Also De Ciccio, "Equilong geometry of third order differential elements," *National Mathematics Magazine*, vol. 19 (1945). Kasner, "Conformal geometry," *Proceedings of the Fifth International Congress of Mathematicians, Cambridge*, vol. 2 (1912), pp. 81-90. See De Ciccio, "Differential geometry in the Kasner plane," *American Mathematical Monthly*, vol. 53, pp. 305-313, where an extensive bibliography on the Conformal geometry of the horn angle is given.

and $v = x - iy$. An arbitrary point transformation T may be written in the form

$$(5) \quad U = U(u, v), \quad V = V(u, v),$$

where the jacobian is

$$(6) \quad J = U_u V_v - U_v V_u \neq 0.$$

The transformation T as defined by (5) is harmonic if and only if each of the components $U(u, v)$ and $V(u, v)$ obey the modified Laplace equation

$$(7) \quad U_{uv} = 0, \quad V_{ur} = 0.$$

The transformation T is conformal if and only if $U_v = V_u = 0$, or $U_u = V_v = 0$.

For convenience, we introduce the letters (p, q, r) to denote the first three total derivatives of v with respect to u . That is, let $p = dv/du$, $q = d^2v/du^2$, $r = d^3v/du^3$. Thus (u, v, p) denotes a differential element of first order, (u, v, p, q) indicates a second order differential element, and (u, v, p, q, r) represents a differential element of third order.

The first extension of the transformation T as defined by (5) is

$$(8) \quad P = \frac{b_0 + b_1 p}{a_0 + a_1 p},$$

where

$$(9) \quad \begin{aligned} a_0 &= U_u, & a_1 &= U_v, & b_0 &= V_u, & b_1 &= V_v, \\ J &= a_0 b_1 - a_1 b_0 = U_u V_v - U_v V_u \neq 0. \end{aligned}$$

Thus the group of arbitrary point transformations induces the general projective group of three-parameters between the pencils of lineal elements whose vertices are at the points corresponding under the transformation T . *This statement is valid even for the infinite set of harmonic transformations.* Of course under the conformal group, equation (8) defines a one-parameter group G_1 , namely, the group of rotations.

The second extension of the transformation T is

$$(10) \quad Q = \frac{c_0 + c_1 p + c_2 p^2 + c_3 p^3 + Jq}{(a_0 + a_1 p)^3},$$

where

$$(11) \quad \begin{aligned} c_0 &= U_u V_{uu} - V_u U_{uu}, & c_1 &= -J_u + 3(U_u V_{uv} - V_u U_{uv}), \\ c_2 &= J_v + 3(U_v V_{uv} - V_v U_{uv}), & c_3 &= U_v V_{vv} - V_v U_{vv}. \end{aligned}$$

The group of arbitrary point transformations induces a group G_8 of eight parameters defined by equations (8) and (10) between the bundles of differ-

ential elements of second order whose vertices are at the points corresponding under the transformation T . It is observed by (11) that *the infinite set of harmonic transformations induces the same eight-parameter group G_8 between the bundles of second order differential elements*. Under the conformal group, equations (8) and (10) define merely a four-parameter group G_4 .

Kasner has presented a complete study of this eight-parameter group G_8 . We shall consider briefly some of his results of which extensions are given in the next section.

A differential element of second order (which passes through a fixed point) is defined by the ordered number pair (p, q) and can be represented as a point in four-dimensional projective space R_4 with homogeneous co-ordinates

$$(12) \quad x_0 : x_1 : x_2 : x_3 : x_4 = 1 : p : p^2 : p^3 : q.$$

The ∞^2 points in R_4 representing the ∞^2 differential elements of second order constitute the manifold defined by

$$(13) \quad x_0x_2 - x_1^2 = 0, \quad x_0x_3 - x_1x_2 = 0, \quad x_1x_3 - x_2^2 = 0.$$

This is a cubic cone S_3 with its vertex at the point $(0, 0, 0, 0, 1)$. Its section in the space $x_4 = 0$ is a twisted cubic.

The geometry of a bundle of differential elements of second order in the plane with respect to the group of arbitrary point transformation is equivalent to projective geometry on a cubic cone S_3 in a space R_4 of four dimensions.

The above principle establishes a classification of differential equations of second order which are algebraic in the derivatives. Kasner defined the *rank n* of such a differential equation of second order to be the degree n of the corresponding algebraic manifold in R_4 . This rank n is invariant under the group of arbitrary point transformations.

In particular, any second order differential equation of the *first rank* is

$$(14) \quad v'' = Av'^3 + Bv'^2 + Cv' + D,$$

where the coefficients (A, B, C, D) are arbitrary functions of (u, v) . This is the cubic type that has been studied extensively by Lie, R. Liouville, Tresse, Kasner, Wilczynski, and Douglas.

The second order differential equations of the *second rank* are

$$(15) \quad 12v''^2 + (B_0 + B_1v' + B_2v'^2 + B_3v'^3)v'' \\ + (C_0 + C_1v' + C_2v'^2 + C_3v'^3 + C_4v'^4 + C_5v'^5 + C_6v'^6) = 0,$$

where the coefficients are arbitrary functions of (u, v) .

Equality of rank is necessary but of course not sufficient for the equivalence of second order differential equations under the group of arbitrary point transformations.

3. Transformation theory of differential elements of third order under the group of arbitrary point transformations. The third extension of the transformation T defined by the general equations (5) is

$$(16) \quad R = \frac{d_0 + d_1p + d_2p^2 + d_3p^3 + d_4p^4 + d_5p^5}{(a_0 + a_1p)^5} \\ + \frac{(h_0 + h_1p + h_2p^2)q - 3a_1Jq^2 + (a_0 + a_1p)Jr}{(a_0 + a_1p)^5},$$

where

$$(17) \quad \begin{aligned} d_0 &= U_u c_{0u} - 3U_{uu}c_0, \\ d_1 &= U_u(c_{0v} + c_{1u}) + U_v c_{0u} - 3U_{uu}c_1 - 6U_{uv}c_0, \\ d_2 &= U_u(c_{1v} + c_{2u}) + U_v(c_{0v} + c_{1u}) - 3U_{uu}c_2 - 6U_{uv}c_1 - 3U_{vv}c_0, \\ d_3 &= U_u(c_{2v} + c_{3u}) + U_v(c_{1v} + c_{2u}) - 3U_{uu}c_3 - 6U_{uv}c_2 - 3U_{vv}c_1, \\ d_4 &= U_u c_{3v} + U_v(c_{2v} + c_{3u}) - 6U_{uv}c_3 - 3U_{vv}c_2, \\ d_5 &= U_v c_{3v} - 3U_{vv}c_3, \\ h_0 &= U_u(c_1 + J_u) - 3JU_{uu} - 3c_0U_v, \\ h_1 &= U_u(J_v + 2c_2) + U_v(J_u - 2c_1) - 6U_{uv}J, \\ h_2 &= 3U_u c_3 + U_v(J_v - c_2) - 3U_{vv}J. \end{aligned}$$

Upon writing the formulas for the partial derivatives c_{0u} , c_{1u} , c_{2v} , c_{3v} , we find that the resulting equations can be solved for U_{uu} , U_{uv} , U_{vv} , U_{vr} . This demonstrates that c_{0u} , c_{1u} , c_{2v} , c_{3v} are independent expressions. By the first, second, fifth, and sixth of equations (17), we find

$$(18) \quad \begin{aligned} c_{0u} &= (1/a_0)(d_0 + 3c_0U_{uu}), & c_{3v} &= (1/a_1)(d_5 + 3c_3U_{vr}), \\ c_{0v} + c_{1u} &= d_1/a_0 - (1/a_0^2)[a_1d_0 + 3(a_1c_0 - a_0c_1)U_{uu} - 6a_0c_0U_{uv}], \\ c_{2v} + c_{3u} &= d_4/a_1 - (1/a_1^2)[a_0d_3 + 3(a_0c_3 - a_1c_2)U_{vr} - 6a_1c_3U_{uv}]. \end{aligned}$$

Substituting these into the third and fourth of equations (17) and then eliminating U_{uu} , U_{vv} , U_{uv} , U_{vr} by means of equations (11), we obtain

$$\begin{aligned}
 & J(a_0^2 d_2 - a_0 a_1 d_1 + a_1^2 d_0) \\
 & \quad - 3(a_0^3 c_0 c_3 + a_0^3 c_1 c_2 - 2a_0^2 a_1 c_0 c_2 - a_0^2 a_1 c_1^2 + 2a_0 a_1^2 c_0 c_1 - a_1^3 c_0^2) \\
 (19) \quad & = a_0^3 [J(c_{2u} + c_{1v}) + 6(a_1 c_1 - a_0 c_2) V_{uv} - 6(b_1 c_1 - b_0 c_2) U_{uv}], \\
 & J(a_1^2 d_3 - a_0 a_1 d_4 + a_0^2 d_5) \\
 & \quad - 3(a_0^3 c_3^2 - 2a_0^2 a_1 c_2 c_3 + a_0 a_1^2 c_2^2 + 2a_0 a_1^2 c_1 c_3 - a_1^3 c_1 c_2 - a_1^3 c_0 c_3) \\
 & = a_1^3 [J(c_{2u} + c_{1v}) + 6(a_1 c_1 - a_0 c_2) V_{uv} - 6(b_1 c_1 - b_0 c_2) U_{uv}].
 \end{aligned}$$

From these equations, we deduce the result

$$\begin{aligned}
 (20) \quad & J[a_0^5 d_5 - a_0^4 a_1 d_4 + a_0^3 a_1^2 d_3 - a_0^2 a_1^3 d_2 + a_0 a_1^4 d_1 - a_1^5 d_0] \\
 & = 3[a_0^3 c_3 - a_0^2 a_1 c_2 + a_0 a_1^2 c_1 - a_1^3 c_0]^2.
 \end{aligned}$$

Upon eliminating U_{uv} , U_{vv} , V_{uu} , V_{vv} by means of equations (11) from the last three of equations (17), we find

$$\begin{aligned}
 (21) \quad & h_0 = -6a_1 c_0 + 3a_0 c_1 - 3a_0(a_0 V_{uv} - b_0 U_{uv}), \\
 & h_1 = 3(a_0 c_2 - a_1 c_1) - 3(a_0 b_1 - a_1 b_0) U_{uv}, \\
 & h_2 = 6a_0 c_3 - 3a_1 c_2 + 3a_1(a_1 V_{uv} - b_1 U_{uv}).
 \end{aligned}$$

From these, we may eliminate the derivatives U_{uv} and V_{uv} . Thus we get the relation

$$(22) \quad a_0^2 h_2 + a_0 a_1 h_1 + a_1^2 h_0 = 6[a_0^3 c_3 - a_0^2 a_1 c_2 + a_0 a_1^2 c_1 - a_1^3 c_0].$$

THEOREM. *The group of arbitrary point transformation induces on the differential elements of third order the group of transformations*

$$\begin{aligned}
 (23) \quad & P = \frac{b_0 + b_1 p}{a_0 + a_1 p}, \quad Q = \frac{c_0 + c_1 p + c_2 p^2 + c_3 p^3 + Jq}{(a_0 + a_1 p)^3}, \\
 & R = \frac{d_0 + d_1 p + d_2 p^2 + d_3 p^3 + d_4 p^4 + d_5 p^5}{(a_0 + a_1 p)^5} \\
 & \quad + \frac{(h_0 + h_1 p + h_2 p^2)q - 3a_1 Jq^2 + (a_0 + a_1 p)Jr}{(a_0 + a_1 p)^5},
 \end{aligned}$$

where the seventeen parameters satisfy the two relationships

$$\begin{aligned}
 (24) \quad & J[a_0^5 d_5 - a_0^4 a_1 d_4 + a_0^3 a_1^2 d_3 - a_0^2 a_1^3 d_2 + a_0 a_1^4 d_1 - a_1^5 d_0] \\
 & = 3[a_0^3 c_3 - a_0^2 a_1 c_2 + a_0 a_1^2 c_1 - a_1^3 c_0]^2, \\
 & a_0^2 h_2 + a_0 a_1 h_1 + a_1^2 h_0 = 6[a_0^3 c_3 - a_0^2 a_1 c_2 + a_0 a_1^2 c_1 - a_1^3 c_0].
 \end{aligned}$$

This induced group is therefore of fifteen essential parameters.

A differential element of third order (which passes through a fixed point) may be denoted by the ordered number triplet (p, q, r) , and can be depicted as a point in eleven-dimensional projective space R_{11} with homogeneous coordinates

$$(25) \quad x_0 : x_1 : x_2 : x_3 : x_4 : x_5 : x_6 : x_7 : x_8 : x_9 : x_{10} : x_{11} \\ = 1 : p : p^2 : p^3 : p^4 : p^5 : q : pq : p^2q : r : pr.$$

The ∞^3 points in R_{11} representing the ∞^3 differential elements of third order form the manifold defined by the system of eight independent equations

$$(26) \quad x_0x_2 - x_1^2 = 0, \quad x_1x_3 - x_2^2 = 0, \quad x_2x_4 - x_3^2 = 0, \quad x_3x_5 - x_4^2 = 0, \\ x_4x_7 - x_5x_6 = 0, \quad x_6x_8 - x_7^2 = 0, \quad x_6x_9 - x_8^2 = 0, \quad x_8x_{11} - x_7x_{10} = 0.$$

This is a three-dimensional algebraic manifold M_5 of the fifth degree.

The geometry of a bundle of ∞^3 differential elements of third order in the plane with respect to the group of arbitrary point transformations is equivalent to projective geometry on a three-dimensional algebraic manifold M_5 of the fifth degree in a space R_{11} of eleven dimensions.

By this principle, we may now define the *rank* n of a differential equation of third order which is algebraic in the derivatives (v', v'', v''') to be the degree n of the corresponding algebraic variety in R_{11} . This rank n is invariant under the group of arbitrary point transformations of the plane of minimal coordinates (u, v) .

In particular, any third order differential equation of the *first rank* is

$$(27) \quad (A_0 + A_1v')v''' + B_0v''^2 + (C_0 + C_1v' + C_2v'^2)v'' \\ + (D_0 + D_1v' + D_2v'^2 + D_3v'^3 + D_4v'^4 + D_5v'^5) = 0.$$

where the twelve coefficients are arbitrary functions of (u, v) . This type includes the differential equation of the ∞^3 Minding geodesic circles of any surface Σ when Σ is represented on the plane in any arbitrary point-to-point fashion,⁴ and also the differential equation of the ∞^3 dynamical trajectories of any positional field of force.⁵

⁴ Kasner and De Cicco, "Families of curves conformally equivalent to circles," *Transactions of the American Mathematical Society*, vol. 49 (1941), pp. 378-391. Also De Cicco, "Equilong maps of the ∞^3 circles," *Transactions of the American Mathematical Society*, vol. 59 (1946), pp. 42-53.

⁵ Kasner, "Differential-geometric aspects of dynamics," *American Mathematical Society Colloquium Publications*, vol. 3, 1913, 1934. Also see a series of papers in *Transactions of the American Mathematical Society*, vols. 7-11 (1906-1910).

Finally see Terracini, "Sobre la equacion diferencial $y''' = G(x, y, y')y'' + H(x, y, y')y'^2$," *Revista de Matematicas de la Universidad de Tucuman (Argentina)*, vol. 2 (1941), pp. 245-329.

4. Transformation theory of differential elements of third order under the infinite set (H) of harmonic transformations. Finally we consider the case of the infinite set (H) of harmonic transformations T . For any such correspondence T , we have $U_{uv} = 0$ and $V_{uv} = 0$. It follows that $c_{2u} + c_{1v} = 0$. Thus by studying equations (19) and (21), we discover the following fundamental proposition.

THEOREM. *The infinite set (H) of harmonic transformations induces on the differential elements of third order the set of transformations (23) where the seventeen parameters satisfy the five relationships*

$$\begin{aligned}
 & J(a_0^2 d_2 - a_0 a_1 d_1 + a_1^2 d_0) \\
 & \quad = 3(a_0^3 c_0 c_3 + a_0^3 c_1 c_2 - 2a_0^2 a_1 c_0 c_2 - a_0^2 a_1 c_1^2 + 2a_0 a_1^2 c_0 c_1 - a_1^3 c_0^2), \\
 (28) \quad & J(a_1^2 d_3 - a_0 a_1 d_4 + a_0^2 d_5) \\
 & \quad = 3(a_0^3 c_3^2 - 2a_0^2 a_1 c_2 c_3 + a_0 a_1^2 c_2^2 + 2a_0 a_1^2 c_1 c_3 - a_1^3 c_1 c_2 - a_1^3 c_0 c_3), \\
 & h_0 = -6a_1 c_0 + 3a_0 c_1, \quad h_1 = 3a_0 c_2 - 3a_1 c_1, \quad h_2 = 6a_0 c_3 - 3a_1 c_2.
 \end{aligned}$$

This induced set S_{12} involves twelve essential parameters.

It is remarked that under the group of conformal transformations, the induced transformations (23) form a six-parameter group G_6 as we have shown in the first article of footnote 3.

Thus the infinite set (H) of harmonic transformations induces between the bundles of differential elements of first, second, and third orders respectively, the following sets of transformations:

- (1). The three-parameter projective group G_3 .
- (2). The eight-parameter group G_8 .
- (3). The twelve-parameter set S_{12} .

In the case of differential elements of the first and second orders, the induced sets are identical with those induced by the larger group of arbitrary point transformations. However, in the theory of differential elements of third order, the infinite set (H) of harmonic transformations induces a set S_{12} of *twelve-parameters*, which is a proper subset of the *fifteen-parameter group* G_{15} induced by the group of all point transformations.

COLUMBIA UNIVERSITY, NEW YORK,
AND
ILLINOIS INSTITUTE OF TECHNOLOGY, CHICAGO.

ON CONVERGENCE OF THE PRODUCT OF BASIC SETS OF POLYNOMIALS.*

By M. N. GHABBOUR.

The subject of basic sets of polynomials has been studied principally by J. M. Whittaker and B. Cannon.¹ For such a set $\{p_n(z)\}$ any polynomial, and in particular the polynomial z^n , admits a unique finite representation of the form

$$(1) \quad z^n = \pi_{n0}p_0(z) + \pi_{n1}p_1(z) + \pi_{n2}p_2(z) + \cdots,$$

Thus any set $\{p_n(z)\}$ of polynomials, where $p_n(z)$ is of degree n , is basic and is said to be *simple*. Generally if the number N_n of nonzero coefficients in (1) is such that $N_n^{1/n}$ tends to 1 as n tends to infinity the set $\{p_n(z)\}$ is called a *Cannon set*. Associated with any function $f(z)$, regular about the origin, there is a *basic series* $\sum \Pi_n f(0)p_n(z)$, where the operators (Π_n) are given by

$$\Pi_n = \pi_{0n} + \pi_{1n}(d/dz) + (1/2!) \pi_{2n}(d^2/dz^2) + \cdots \quad (n = 0, 1, 2, \cdots).$$

This series is said to *represent* $f(z)$ in $|z| \leq R$, where the function is regular, if it converges uniformly to $f(z)$ in $|z| \leq R$. The basic series of the set $\{p_n(z)\}$ is said to be *effective* in $|z| \leq R$, if it represents, in $|z| \leq R$, every function which is regular there. This is equivalent to the simpler statement that the set $\{p_n(z)\}$ is *effective* in $|z| \leq R$. Let

$$(2) \quad \delta_n(R) = \sum_i |\pi_{ni}| A_i(R),$$

where $A_n(R)$ is the maximum modulus of $p_n(z)$ in $|z| \leq R$. Cannon (1) has² shewn that a necessary and sufficient condition for a Cannon set $\{p_n(z)\}$ to be effective in $|z| \leq R$, is that

$$(3) \quad \delta(R) = \overline{\lim}_{n \rightarrow \infty} \{\delta_n(R)\}^{1/n} = R,$$

for any value of $R > 0$. Suppose that $p_n(z) = p_{n0} + p_{n1}z + p_{n2}z^2 + \cdots$.

* Received November 21, 1946.

¹ See the list of references at the end.

² The numbers beside the authors refer to the order of their references as given in the list.

The matrix $P = (p_{ij})$ is called the *matrix of coefficients* of the set $\{p_n(z)\}$. If P_1 and P_2 are the matrices of coefficients of the respective basic sets $\{p_n(z)\}$ and $\{q_n(z)\}$, then it has been shewn by Whittaker (4) that the matrix $P_1 \cdot P_2$ will be the matrix of coefficients of the basic set $\{u_n(z)\}$ given by

$$(4) \quad u_n(z) = p_{n0}q_0(z) + p_{n1}q_1(z) + p_{n2}q_2(z) + \dots$$

The set $\{u_n(z)\}$ is defined to be the *product* of the sets $\{p_n(z)\}$ and $\{q_n(z)\}$ in the given order. It was proposed by Professor Whittaker to study the convergence of the product set, or more precisely, to find out when the product set $\{u_n(z)\}$ is effective. In this note the problem is considered for the case where $\{p_n(z)\}$ is a Cannon set and $\{q_n(z)\}$ is a set for which $\lim_{n \rightarrow \infty} D_n/n$ is finite, where D_n is defined by Whittaker (4) to be the degree of the polynomial of highest degree in (1).

Let $\{p_n(z)\}$ and $\{q_n(z)\}$ be basic sets of polynomials; then the polynomial $p_n(z)$ can be uniquely expressed as a finite linear combination of the polynomials $\{q_n(z)\}$ in the form

$$(5) \quad p_n(z) = \sum_i \tilde{\omega}_{ni} q_i(z).$$

Suppose that $f(z)$ is any function that can be represented by the basic series of $\{p_n(z)\}$ in some circle about the origin, in the form $f(z) = \sum_{n=0}^{\infty} a_n p_n(z)$. Associated with such a function there is, therefore, a series $\sum c_n q_n(z)$, where

$$(6) \quad c_n = a_0 \tilde{\omega}_{0n} + a_1 \tilde{\omega}_{1n} + a_2 \tilde{\omega}_{2n} + \dots$$

This series represents $f(z)$ in $|z| \leq R$, if it converges uniformly to $f(z)$ in $|z| \leq R$. The number N_n of nonzero coefficients in (5) is assumed to be such that

$$(7) \quad N_n^{1/n} \rightarrow 1, \text{ as } n \rightarrow \infty.$$

Write $\phi_n(R) = \sum_i |\tilde{\omega}_{ni}| B_i(R)$, where $B_n(R)$ is the maximum modulus of $q_n(z)$ in $|z| \leq R$. I shall prove the following

LEMMA. *Let $\{p_n(z)\}$ be a basic set of polynomials and suppose that the basic set $\{q_n(z)\}$ satisfies (7), and that for any value of $R > 0$, $\lim_{n \rightarrow \infty} \{\phi_n(R)/A_n(R)\}^{1/n} > 1$; then there is a function $f(z)$, which is represented by the basic series of $\{p_n(z)\}$ in $|z| \leq R$, and yet the series $\sum c_n q_n(z)$ does not represent it in $|z| \leq R$.³*

³ The proof of the lemma is similar to the proof of the above theorem of Cannon (1).

For, let

$$(8) \quad \overline{\lim}_{n \rightarrow \infty} \{\phi_n(R)/A_n(R)\}^{1/n} \geq \alpha > 1.$$

Applying Hadamard's three circles theorem we can easily construct the sequence of numbers (R_n) such that

$$(9) \quad 1/\alpha_1 \leq \{A_n(R)/A_n(R_n)\}^{1/n} \leq 1/\alpha_2, \quad \text{when } 1 < \alpha_2 < \alpha_1 < \alpha.$$

Combining (8) and (9) we obtain

$$(10) \quad \overline{\lim}_{n \rightarrow \infty} \{\phi_n(R)/A_n(R_n)\}^{1/n} > 1.$$

It follows that the series $\sum |\tilde{\omega}_{ns}|/A_n(R_n)$ may

(a) either diverge for at least one value of s , s_1 say, and in this case the required function will be the function $f(z) = \sum_{n=0}^{\infty} p_n(z) e^{-i\beta_{ns_1}}/A_n(R_n)$, where $\beta_{ns_1} = \arg \tilde{\omega}_{ns_1}$, or

(b) converge for all values of s . Here applying (7) and (10) we construct an infinite sequence of pairs of integers (m_r, q_r) such that $m_{r+1} > m_r$, $q_{r+1} > q_r$ and

$$(11) \quad |\tilde{\omega}_{m_i, q_i}| B_{q_r}(R) > A_{m_r}(R_{m_r}).$$

Now by the convergence of the series $\sum |\tilde{\omega}_{ns}|/A_n(R_n)$ and the construction of the matrix $(\tilde{\omega}_{ij})$ we construct, in a successive manner, a sequence of integers (λ_k) and a subsequence $(\tilde{\omega}_{b_k d_k})$ out of the sequence $(\tilde{\omega}_{m_r q_r})$ such that $\tilde{\omega}_{b_1 d_1} = \tilde{\omega}_{m_1 q_1}$, $b_k > \lambda_{k-1}$, $\tilde{\omega}_{n d_k} = 0$ for $n < b_k$, and $\sum_{n=\lambda_k}^{\infty} |\tilde{\omega}_{n d_k}|/A_n(R_n) < \frac{1}{2} |\tilde{\omega}_{b_k d_k}|/A_{b_k}(R_{b_k})$. The required function in this case will be the function $f(z) = \sum a_n p_n(z)$, for which $a_n = 1/A_n(R_n)$ when $n = b_1, b_2, b_3, \dots$, and $a_n = 0$ otherwise.

We now consider the main problem and set

$$(12) \quad z^n = \sum_i \lambda_{ni} q_i(z).$$

Then from (1) and (12) and the definition of the product set it can be easily shewn that

$$(13) \quad z^n = \sum_i u_i(z) \sum_r \lambda_{ni} \pi_{ri},$$

and ⁴

⁴ The relation (14) attributes an interesting meaning to the concept of product sets. Let $f(z) = \sum a_n z^n$ be represented in a series of polynomials of the form $\sum c_n p_n(z)$; then

$$(14) \quad q_n(z) = \sum_i \pi_{ni} u_i(z).$$

Write

$$(15) \quad \psi_n(R) = \sum_i |\pi_{ni}| M_i(R),$$

where $M_n(R)$ is the maximum modulus of $u_n(z)$ in $|z| \leq R$. For the sake of uniformity we shall write $\delta_n^{(1)}(R)$ for $\delta_n(R)$ and $\delta^{(1)}(R)$ for $\delta(R)$ as given by (2) and (3) and $\delta_n^{(2)}(R)$ and $\delta^{(2)}(R)$ and $\Delta_n(R)$ and $\Delta(R)$ for the corresponding expressions for the sets $\{q_n(z)\}$ and $\{u_n(z)\}$, respectively. We first prove the following

THEOREM I. *Let $\{p_n(z)\}$ be a Cannon set and let the set $\{q_n(z)\}$ be subject to the condition that $\lim_{n \rightarrow \infty} D_n/n$ is finite and suppose that $\{q_n(z)\}$ is effective in $|z| \leq R$. Then the necessary and sufficient condition for the product set $\{u_n(z)\}$ to be effective in $|z| \leq R$ is that*

$$(16) \quad \overline{\lim}_{n \rightarrow \infty} \{\psi_n(R)/B_n(R)\}^{1/n} = 1.$$

The necessity follows from the above lemma, noting that when $\{p_n(z)\}$ is a Cannon set then the expression in (14) will satisfy (7). Also if both the sets $\{u_n(z)\}$ and $\{q_n(z)\}$ are effective in $|z| \leq R$, then the basic series of $\{u_n(z)\}$ will represent, in $|z| \leq R$, every function represented by the basic series of $\{q_n(z)\}$ in $|z| \leq R$. Applying the above lemma, (16) follows at once.

To prove the sufficiency of (16), a positive number K is chosen for the set $\{q_n(z)\}$ such that

$$(17) \quad D_n/n < K, \text{ for all values of } n > n_0, \text{ say.}$$

Now from (13) we obtain

$$\Delta_n(R) < \sum_r |\lambda_{nr}| \sum_i |\pi_{ri}| M_i(R) < \delta_n^{(2)}(R) \psi_{t_n}(R)/B_{t_n}(R),$$

where $\psi_{t_n}(R)/B_{t_n}(R) = \max_r \psi_r(R)/B_r(R) \geq 1$. If t_n remains finite then

$$(18) \quad \Delta(R) = \overline{\lim}_{n \rightarrow \infty} \{\Delta_n(R)\}^{1/n} \leq \delta^{(2)}(R).$$

If, on the other hand, t_n tends to infinity with n , then in view of (17) we have, for $n > n_0$,

according to (14) a representation of $\Sigma a_n q_n(z)$ will be $\Sigma c_n u_n(z)$ and it is seen that the last two series are derived from the former two series by putting in each $q_n(z)$ instead of z^n . The question of product sets is therefore equivalent to that of finding the conditions governing the series $\Sigma c_n p_n(z)$ associated with the function $\Sigma a_n z^n$ in order that the series $\Sigma c_n u_n(z)$ should represent the function $\Sigma a_n q_n(z)$.

$$\{\Delta_n(R)\}^{1/n} < \{\delta_n^{(2)}(R)\}^{1/n} \cdot \{\psi_{t_n}(R)/B_{t_n}(R)\}^{k/t_n},$$

and applying (16), (18) follows also. Since the set $\{q_n(z)\}$ is effective in $|z| \leq R$, then according to (3), $\delta^{(2)}(R) = R$, which ensures that the set $\{u_n(z)\}$ is also effective in $|z| \leq R$, as required.

Now suppose that $\lim_{n \rightarrow \infty} \{B_n(R)/R^n\}^{1/n}$ exists and is finite. In fact let

$$(19) \quad \lim_{i \rightarrow \infty} \{B_n(R)/R^n\}^{1/n} = a < \infty.$$

The following theorem defines completely the region where the product set $\{u_n(z)\}$ is effective.

THEOREM II. *Let $\{p_n(z)\}$ be a Cannon set and let the set $\{q_n(z)\}$ be such that $\lim_{i \rightarrow \infty} D_n/n$ is finite, and suppose that $\{q_n(z)\}$ is effective in $|z| \leq R$, where it satisfies (19), and $\delta^{(1)}(aR)$ is continuous. Then the product set $\{u_n(z)\}$ will be effective in $|z| \leq R$, if, and only if, $\{p_n(z)\}$ is effective in $|z| \leq aR$.*

For let $\rho = aR$ and choose the numbers R_1, R_2, R_3 and R_4 such that

$$(20) \quad R_1 < R_2 < \rho < R_3 < R_4.$$

Then according to (19) there exists a positive integer n_0 such that

$$(21) \quad R_2^n < B_n(R) < R_3^n,$$

for $n > n_0$. Hence from (4), (20) and (21) we get

$$(22) \quad \begin{aligned} M_n(R) &< \left\{ \sum_{r \leq n_0} + \sum_{r > n_0} \right\} |p_{nr}| B_r(R) \\ &< (n_0 + 1) B_s(R) \cdot A_n(R_4)/R_4^s + A_n(R_4) \sum_{r=0}^{\infty} (R_3/R_4)^r \\ &< A_n(R_4) \cdot S(R, R_3, R_4), \end{aligned}$$

where $B_s(R)/R_4^s = \max_{r \leq n_0} B_r(R)/R_4^r$, so that $S(R, R_3, R_4)$ is finite. In a similar manner it can be shown, by the aid of (20) and (21), that

$$(23) \quad A_n(R_1) < M_n(R)/T(R, R_1, R_2),$$

where $T(R, R_1, R_2)$ is positive and finite. Combining (22) and (23) together we get, in view of (15) and (21),

$$T(R, R_1, R_2) \cdot \delta_n^{(1)}(R_1)/R_3^n < \psi_n(R)/B_n(R) < \delta_n^{(1)}(R_4) \cdot S(R, R_3, R_4)/R_2^n.$$

Since R_1 and R_4 can be taken as near to ρ as we please and since $\delta^{(1)}(\rho)$ is continuous, then, making n tend to infinity, we conclude that

$$(24) \quad \overline{\lim}_{i \rightarrow \infty} \{\psi_n(R)/B_n(R)\}^{1/n} = \delta^{(1)}(\rho)/\rho.$$

The required result now follows by appealing to the above theorem of Cannon and to Theorem I.

The following corollaries are direct applications of Theorem II for some important special cases:

COROLLARY I. *Let $\{p_n(z)\}$ be a Cannon set and let the set $\{q_n(z)\}$ be such that $\overline{\lim}_{n \rightarrow \infty} D_n/n$ is finite. Suppose that the set $\{q_n(z)\}$ is effective in $|z| \leq R$, where $\lim_{n \rightarrow \infty} \{B_n(R)/R^n\}^{1/n} = 1$, and $\delta^{(1)}(R)$ is continuous. Then the product set $\{u_n(z)\}$ is effective in $|z| \leq R$, if and only if, $\{p_n(z)\}$ is effective in $|z| \leq R$.⁵*

COROLLARY II. *Let $\{p_n(z)\}$ be a Cannon set and let $\{q_n(z)\}$ be a simple set in which the coefficient of z^n in $q_n(z)$ is unity. Suppose that $\{q_n(z)\}$ is effective in $|z| \leq R$, where $\delta^{(1)}(R)$ is continuous. Then the product set $\{u_n(z)\}$ is effective in $|z| \leq R$, if, and only if, $\{p_n(z)\}$ is effective in $|z| \leq R$.*

The result follows from Corollary I, if we note that for such a simple set $\{q_n(z)\}$, $\lim_{i \rightarrow \infty} \{B_n(R)/R^n\}^{1/n} = 1$.

COROLLARY III. *Let $\{p_n(z)\}$ be a set for which $\lim_{n \rightarrow \infty} D_n/n = 1$, and let the set $\{q_n(z)\}$ be such that $\overline{\lim}_{n \rightarrow \infty} D_n/n$ is finite and suppose that $\{q_n(z)\}$ is effective in $|z| \leq R$, where $\lim_{n \rightarrow \infty} \{B_n(R)/R^n\}^{1/n} \geq 1$. Then the product set $\{u_n(z)\}$ is effective in $|z| \leq R$, if $\{p_n(z)\}$ is effective in $|z| \leq R$.*

For, according to Whittaker (4), $\delta^{(1)}(R)/R < \delta^{(1)}(r)/r$, for $r < R$. Combining this with (24) the required result is obtained.

It will be observed that Theorem II gives us instances for the case in which only the 'inner' set $\{q_n(z)\}$ is effective and yet the product set is effective in the same circle. It also affords us with instances for the case where both the sets $\{p_n(z)\}$ and $\{q_n(z)\}$ are effective in $|z| \leq R$, and yet the product set is not effective in $|z| \leq R$. These instances are illustrated in the following example:

Example I. Consider the basic sets $\{p_n(z)\}$ and $\{q_n(z)\}$ given by $p_n(z) = 1 + z^n + z^{2n}/4^n$, when n is odd, and $p_n(z) = z^n$ when n is even, and $q_n(z) = z^n + 2^n z^{n+1}$ when n is odd, and $q_n(z) = 2^n z^{n-1} + z^n$ when $n > 0$ is even and $q_0(z) = 1$.

⁵ The condition that $\lim_{n \rightarrow \infty} \{B_n(R)/R^n\}^{1/n} = 1$ has been obtained as a sufficient condition for the effectiveness of $\{u_n(z)\}$ if both $\{p_n(z)\}$ and $\{q_n(z)\}$ are effective.

It is easily seen that $\{p_n(z)\}$ is effective in $|z| \leq R$ for the values of R for which $1 \leq R \leq 4$, and that $\{q_n(z)\}$ is effective in the whole plane. Also it is noted that $\lim_{n \rightarrow \infty} \{B_n(R)/R^n\}^{1/n} = 2$. Forming the product set $\{u_n(z)\}$ of the sets $\{p_n(z)\}$ and $\{q_n(z)\}$ we get

$$u_n(z) = 1 + z^n + 2^n z^{n+1} + z^{2n-1} + z^{2n}/4^n, \text{ when } n \text{ is odd, and}$$

$$u_n(z) = 2^n z^{n-1} + z^n, \text{ when } n > 0 \text{ is even, and } u_0(z) = 1.$$

It can be easily shown that $\overline{\lim}_{n \rightarrow \infty} \{\psi_n(R)/B_n(R)\}^{1/n} = 1$ when $\frac{1}{2} \leq R \leq 2$ and thus, by Theorem I, the set $\{u_n(z)\}$ is effective in $|z| \leq R$ for the values of R for which $\frac{1}{2} \leq R \leq 2$, as is expected from Theorem II. Also we note that in the interval $\frac{1}{2} \leq R < 1$, the set $\{q_n(z)\}$ is effective in $|z| \leq R$, while the set $\{p_n(z)\}$ is not effective and yet the product set $\{u_n(z)\}$ is effective. Also in the interval $2 < R \leq 4$, both the sets $\{p_n(z)\}$ and $\{q_n(z)\}$ are effective in $|z| \leq R$, and yet the product set $\{u_n(z)\}$ is not effective.

Now the condition that $\lim_{n \rightarrow \infty} \{B_n(R)/R^n\}^{1/n}$ exists and is finite, is necessary for the truth of Theorem II in the sense that if this condition is not satisfied then the theorem is no longer true. In the first place if $\lim_{n \rightarrow \infty} \{B_n(R)/R^n\}^{1/n}$ is infinite, then instances can be found in which both the sets $\{p_n(z)\}$ and $\{q_n(z)\}$ are effective in the whole plane while the product set $\{u_n(z)\}$ is nowhere effective.⁶

Moreover, if

$$(25) \quad 0 < a = \lim_{n \rightarrow \infty} \{B_n(R)/R^n\}^{1/n} < \overline{\lim}_{n \rightarrow \infty} \{B_n(R)/R^n\}^{1/n} = b \leq \infty,$$

then Theorem II no longer holds, as is seen from the following

⁶ Thus consider, for example, the basic sets $\{p_n(z)\}$ and $\{q_n(z)\}$ given by

$$p_n(z) = zn + z^{2n}/n^n, \text{ when } n \text{ is odd, and } p_n(z) = z^n,$$

when n is even, and

$$q_n(z) = z^n + n^n z^{n+1}, \text{ when } n \text{ is odd, and } q_n(z) = (n-1)^{2(n-1)} z^{n-1} + z^n$$

when $n > 0$ is even, and $q_0(z) = 1$.

Accordingly the product set $\{u_n(z)\}$ will be given by

$$u_0(z) = 1, \text{ and } u_n(z) = (n-1)^{2(n-1)} z^{n-1} + z^n,$$

when $n > 0$ is even, and

$$u_n(z) = z^n + n^n z^{n+1} + (2n-1)^{2(2n-1)} z^{2n-1}/n^n + z^{2n}/n^n,$$

when n is odd.

It is easily seen that $\{B_n(R)/R^n\}^{1/n}$ tends to infinity with n , that both $\{p_n(z)\}$ and $\{q_n(z)\}$ are effective in the whole plane, and finally that $\overline{\lim}_{n \rightarrow \infty} \{\psi_n(R)/B_n(R)\}^{1/n} = \infty$ for all values of $R > 0$.

THEOREM III. Let $\{q_n(z)\}$ be a basic set for which $\overline{\lim_{n \rightarrow \infty}} D_n/n$ is finite and suppose that $\{q_n(z)\}$ is effective in $|z| \leq R$, where it satisfies (25). Then there exists a Cannon set $\{p_n(z)\}$ which is effective in $|z| \leq r$, for all positive values of $r \leq bR$, such that the product set $\{u_n(z)\}$ is not effective in $|z| \leq R$.

For, write $aR = \rho$, then, given any number $R_1 > \rho$, there exists an infinite sequence of positive integers (l_r) such that

$$(26) \quad B_{l_r}(R) < R_1^{l_r} \quad (r = 1, 2, 3, \dots).$$

Writing $\{B_n(R)/R^n\}^{1/n} = b_n$, then there exists another finite sequence of positive integers (m_r) , having no common members with (l_r) such that

$$(27) \quad b_{m_r} \rightarrow b, \text{ as } r \rightarrow \infty.$$

We form the infinite sequence of triads of integers $(\lambda_i, \mu_i, \nu_i)$ for $i = 1, 2, 3, \dots$, where

$$(28) \quad \begin{cases} (\lambda_i) \text{ belongs to } (l_r) \text{ and } (\mu_i) \text{ belongs to } (m_r), \text{ and} \\ \lambda_i < \mu_i \text{ and } \nu_i = \mu_i - \lambda_i. \end{cases}$$

The required Cannon set $\{p_n(z)\}$ is given by

$$(29) \quad p_{\lambda_i}(z) = z^{\lambda_i} + (z^{\mu_i}/(b_{\mu_i} \cdot R)^{\nu_i}) \text{ and } p_n(z) = z^n, \text{ otherwise.}$$

It can be easily seen, using (27), (28) and (29), that the set $\{p_n(z)\}$ is either effective in the whole plane if $\lim_{i \rightarrow \infty} (\mu_i/\lambda_i) = 1$, or else is effective in $|z| \leq r$, for all positive values of $r \leq bR$. Now the product set $\{u_n(z)\}$ is given by

$$u_{\lambda_i}(z) = q_{\lambda_i}(z) + q_{\mu_i}(z)/(b_{\mu_i} \cdot R)^{\nu_i}, \text{ and } u_n(z) = z^n, \text{ otherwise.}$$

Hence in view of (26) and (28) we have

$$\frac{\psi_{\lambda_i}(R)}{B_{\lambda_i}(R)} = \frac{M_{\lambda_i}(R) + (1/(b_{\mu_i} \cdot R)^{\nu_i})M_{\mu_i}(R)}{B_{\lambda_i}(R)} > \left(\frac{b_{\mu_i} \cdot R}{R_1}\right)^{\lambda_i}.$$

Allowing i to tend to infinity we conclude, according to (27), that

$$\overline{\lim_{n \rightarrow \infty}} \{\psi_n(R)/B_n(R)\}^{1/n} \geq \lim_{i \rightarrow \infty} \{\psi_{\lambda_i}(R)/B_{\lambda_i}(R)\}^{1/\lambda_i} \geq bR/R_1 > 1,$$

since we can take R_1 as near to ρ as we please. Hence by Theorem I the set $\{u_n(z)\}$ is not effective in $|z| \leq R$, as required.

I append the following illustrative example:

Example II. Consider the basic set $\{q_n(z)\}$ given by

$$\begin{aligned}
 q_n(z) &= z^n + 2^n z^{n+2} && \text{when } n = 3h, \text{ and} \\
 q_n(z) &= 2^n z^{n-1} + z^n + 2^n z^{n+1} && \text{when } n = 3h + 1, \text{ and} \\
 q_n(z) &= z^n && \text{when } n = 3h + 2,
 \end{aligned}$$

for $h = 0, 1, 2, \dots$. It can be easily shewn that $\{q_n(z)\}$ is effective in the whole plane, and

$$1 = \lim_{n \rightarrow \infty} \{B_n(R)/R^n\}^{1/n} < \overline{\lim}_{n \rightarrow \infty} \{B_n(R)/R^n\}^{1/n} = 2.$$

Forming the product set $\{u_n(z)\}$ of the sets $\{p_n(z)\}$ as given in Example I and $\{q_n(z)\}$ as given above, we get

$$\begin{aligned}
 u_n(z) &= z^n + 2^n z^{n+1}; && \text{for } n = 6h, \\
 u_n(z) &= 1 + z + 2^n z^{n-1} + z^n + 2^n z^{n+1} + z^{2n}/4^n; && \text{for } n = 6h + 1, \\
 u_n(z) &= z^n; && \text{for } n = 6h + 2, \\
 u_n(z) &= 1 + z + z^n + 2^n z^{n+1} + z^{2n}/4^n + z^{2n+1}; && \text{for } n = 6h + 3, \\
 u_n(z) &= 2^n z^{n-1} + z^n + 2^n z^{n+1}; && \text{for } n = 6h + 4, \\
 u_n(z) &= 1 + z + z^n + z^{2n-1} + z^{2n}/4^n + z^{2n+1}; && \text{for } n = 6h + 5,
 \end{aligned}$$

$h = 0, 1, 2, \dots$. It is easily seen that $\overline{\lim}_{n \rightarrow \infty} \{\psi_n(R)/B_n(R)\}^{1/n} = 1$ only when $R = 1$ and is otherwise greater than 1, so that the set $\{u_n(z)\}$ is only effective in $|z| \leq 1$.

FAROUK I UNIVERSITY,
ALEXANDRIA, EGYPT.

REFERENCES

J. M. Wittaker:

1. "On series of polynomials," *Quarterly Journal of Mathematics*, vol. 5 (1934), p. 224.
2. "On the uniqueness of expansion in polynomials," *Journal of the London Mathematical Society*, vol. 10 (1935), p. 108.
3. *Interpolatory Function Theory*, Cambridge (1935).
4. *Series of Polynomials*, Tract of Fouad I University (1943).

B. Cannon:

1. "On the convergence of series of polynomials," *Proceedings of the London Mathematical Society* (2), vol. 43 (1937), p. 348.
2. "On the convergence properties of basic series," *Journal of the London Mathematical Society*, vol. 14 (1939), p. 51.
3. "On the representation of integral functions by general basic series," *Mathematische Zeitschrift*, vol. 4 (1939), p. 185.

ELIMINANTS.*

By JOSEPH MILLER THOMAS.

From the theory of functional dependence is known the existence of polynomials $F(x, y)$ which vanish identically in t when the indeterminates x, y are replaced by given polynomials in the single indeterminate t . A specific polynomial $E(x, y)$ of this sort is defined in the present paper (1) as the eliminant of two given polynomials. This polynomial has been employed by Perron [1; p. 223]. Taking the complex numbers for coefficient field (the extension to fields of characteristic zero is immediate), here we develop certain properties of this eliminant. Firstly, the eliminant is the k -th power of an irreducible polynomial f . Secondly, the eliminant is reducible (that is, $1 < k$) if and only if the two polynomials $x(t), y(t)$ are also polynomials in a second indeterminate u which itself is a polynomial of degree k in the original indeterminate t . Thirdly, whereas elimination of t by the eliminant gives Af^k , where A is a non-zero complex number, elimination by the usual division process gives hf , where h is an extraneous factor, whose nature is precisely stated, namely, h is the product of non-negative integral powers of the initials appearing in the division sequence and h may involve x and y . Fourthly, the eliminant is used to give algebraic conditions that a single polynomial $y(t)$ be a polynomial in a second indeterminate u whose degree satisfies $1 < \deg_t u < \deg_t y$. An elegant solution of the last problem (for a single polynomial and a parameter of any degree) from group-theoretic and function-theoretic considerations has previously been given by Ritt [2].

1. The eliminant. The parametric equations

$$(1.1) \quad x = a_0 t^m + \cdots + a_m, \quad y = b_0 t^n + \cdots + b_n, \quad a_0 b_0 \neq 0$$

define a curve which is the locus of the point (x, y) . Let the resultant of the two polynomials be denoted by $R(a_m, b_n)$, a notation which puts in evidence its dependence on the constant terms of the polynomials. Put

$$(1.2) \quad E(x, y) = R(a_m - x, b_n - y).$$

* Received July 26, 1946; presented to the American Mathematical Society, November 2, 1946.

The equation

$$(1.3) \quad E(x, y) = 0$$

is a necessary and sufficient condition that for a given (x, y) there exist a t satisfying (1.1). Hence (1.3) is the equation of the curve. Its left member will be called the *eliminant* of the two polynomials. If the values (1.1) are introduced into E , then

$$(1.4) \quad E(x, y) \equiv 0$$

in t . The eliminant satisfies

$$(1.5) \quad \deg_{x,y} E = \max(m, n), \quad \deg_x E = n, \quad \deg_y E = m;$$

all of these relations are independent of the values given (x, y) .

2. Reducibility. Let the factorization of E into irreducible factors be

$$E(x, y) = f(x, y)g(x, y) \cdots$$

Under the substitution (1.1) at least one of the factors, say $f(x, y)$, becomes identically zero. If (x, y) is a root of $g(x, y)$, there exists a t satisfying (1.1). Hence (x, y) is also a root of $f(x, y)$, f and g are associates and

$$(2.1) \quad E(x, y) = Af^k,$$

where A is a non-zero complex number, f is an irreducible polynomial in x, y and k is a positive integer. The curve is got by imagining the points of the unicursal curve $f(x, y) = 0$ traced k times. From (2.1) we have

$$(2.2) \quad m = k \deg_y f, \quad n = k \deg_x f.$$

Suppose that x, y are also polynomials in a parameter u , which itself is a polynomial in t :

$$(2.3) \quad u = c_0 t^p + c_1 t^{p-1} + \cdots + c_p, \quad c_0 \neq 0.$$

Denote by $E(x, y; u)$ the E formed with respect to u , and by $E(x, y; t)$ that previously considered. Since the point set constituting the solution of $E(x, y) = 0$ is the same for both parameters, f can be taken the same for u as for t . Hence $E(x, y; u) = Bf^l$, and

$$(2.4) \quad E(x, y; t) = C[E(x, y; u)]^p, \quad k = pl.$$

The parameter t will be called *reducible* or *irreducible* according as there does or does not exist for $1 < p$ a polynomial u in terms of which x and y can be simultaneously expressed. From (2.4) it follows that if the parameter t is reducible, so also is the eliminant $E(x, y; t)$. The converse of this proposition is also true as will be seen in 4.

3. Highest common factor. Generalizing (1.2) put

$$E_i(x, y) = R_i(a_m - x, b_n - y),$$

where R_i is the i -th subresultant and $E_0 = E$. Let a positive integer l be defined by

$$(3.1) \quad E_0(x, y) \equiv 0, \dots, E_{l-1}(x, y) \equiv 0, E_l(x, y) \not\equiv 0,$$

the sign \equiv being read "equals identically in t ." The curves $E_0 = 0$, $E_k = 0$ intersect in a finite number of points, which are singular on $f = 0$. For an ordinary (x, y) the highest common factor of

$$(3.2) \quad r_0 = a_0 t^m + \dots + a_m - x, \quad r_1 = b_0 t^n + \dots + b_n - y$$

has degree l and is given by

$$(3.3) \quad E_l(x, y)t^l + \dots,$$

where the unwritten terms are easily specified polynomials in (x, y) . The locus of points at which the discriminant of (3.3) vanishes intersects $E(x, y) = 0$ in a finite number of points. Except at these and singular points, (3.3) has exactly l distinct roots t_1, \dots, t_l all of which give the same point (x, y) on $f = 0$. That the roots of (3.3) may not all be distinct at an ordinary point is shown by $x = t^2$, $y = t^2$, $E(x, y) = -(x - y)^2$.

It is now convenient to suppose the notation chosen so that $n \leq m$ and consequently $\deg_{x,y} E = m$. By avoiding a finite number of points, it is possible to choose a point (x, y) such that $E_l(x, y) \not\equiv 0$ and such that the line $X = x$ meets the curve $f = 0$ in m/k distinct points. The m parameter values of these points fall into m/k sets of l each. Hence $k = l$ and the positive integer k in (2.1) can be determined by

$$(3.4) \quad E_0(x, y) \equiv 0, \dots, E_{k-1}(x, y) \equiv 0, E_k(x, y) \not\equiv 0.$$

4. The fundamental theorem. If t and s are indeterminates, the polynomial $x(t) - x(s)$ is divisible by $t - s$. If it is written as the product of two factors, the factorization can be given the form

$$(4.1) \quad x(t) - x(s) = (t^k + \cdots - s^k)(a_0 t^{m-k} + \cdots + a_0 s^{m-k}),$$

the first factor being divisible by $t - s$. The highest common factor of $x(t) - x(s)$ and $y(t) - y(s)$ can therefore be assumed in the form

$$u(t, s) = t^k + \cdots - s^k.$$

Let the polynomials $u(t, s_1)$ and $u(t_1, s)$ have the roots s_1, \cdots, s_k and t_1, \cdots, t_k respectively. The parameter values s_1, \cdots, s_k all give the same point, say (x, y) and the t_1, \cdots, t_k the same point, say (x_0, y_0) . Write

$$(4.2) \quad u(t, s) = t^k + \alpha_1(s)t^{k-1} + \cdots + \alpha_k(s),$$

where $\alpha_1, \cdots, \alpha_{k-1}$ are polynomials of degree at most $k-1$ and α_k is of degree k . Consider an ordinary point s_1 . Since (4.2) is an associate of (3.3) its coefficients depend only on (x, y) and

$$\alpha_j(s_1) = \cdots = \alpha_j(s_k) \quad (j = 1, 2, \cdots, k).$$

The equations

$$\alpha_j(s) - \alpha_j(s_1) = 0 \quad (j = 1, 2, \cdots, k-1)$$

accordingly are identities. Moreover the polynomial $\alpha_k(s) - \alpha_k(s_1)$ has the roots s_1, \cdots, s_k and therefore is $-u(s, s_1)$. Writing $u(t)$ for $u(t, s_1)$ we therefore have

$$(4.3) \quad u(t, s) = u(t) - u(s).$$

Let $u(t)$ be fixed by assigning an s_1 corresponding to an ordinary point. The k parameter values belonging to each point (x, y) on $E(x, y) = 0$ also belong to some point (x, y) on the curve

$$(4.4) \quad x = x(t), \quad y = u(t).$$

The eliminant of (4.4) is therefore the k -th power of an irreducible polynomial of degree 1 in x . Hence x is a polynomial in u . A similar argument applies to y and we have

THEOREM 4.1. *The eliminant is reducible if and only if the parameter is reducible.*

The reducibility of the eliminant is equivalent to the reducibility of the resultant in which the two coefficients a_m, b_n (and only those) are indeterminates.

Once it has been established that there are k parameter values giving each point, Lüroth's theorem shows that x, y are rational functions (not polynomials) in a reduced parameter and that the reduced parameter is a rational function of x, y . That the second result cannot be strengthened in the present case is shown by the example:

$$x = t^6, \quad y = t^4, \quad E(x, y) = (x^2 - y^3)^2, \quad u(t) = t^2 - s_1^2 = (x/y) - (x_0/y_0).$$

If u were a polynomial in (x, y) , we should have identically

$$x - yg(x, y) = h(x, y)(x^2 - y^3),$$

where g, h are polynomials. Evaluation for $y = 0$ gives the contradiction $x = x^2h(x, 0)$.

5. Connection with division process. Still under the assumption $n \leq m$, let r_2 be the remainder when $b_0^{m-n+1}r_0$ is divided by r_1 for indeterminate x, y (see (3.2) for the definition of r_0, r_1). In this way form a division sequence

$$(5.1) \quad r_0, r_1, r_2, \dots, r$$

whose last term r is not zero and does not contain t . The polynomials in (5.1) are equal to the corresponding polynomials in the division sequence formed for the irreducible parameter, the only possible change being in the parameter. The degree differences for the original parameter will be k times those for the irreducible parameter.

By reference to [3; § 6] it is seen that $E_j(x, y)$ vanishes identically in x, y if j is not divisible by k . This can be used to modify (3.4).

It can be shown [3; (4.8)] that

$$(5.2) \quad \pm g(x, y)R_0(a_m - x, b_n - y) = I^d,$$

where $g(x, y)$ is the product of non-negative integral powers of initials of polynomials which precede $r(x, y)$ in the sequence (5.1), where I is the initial of $r(x, y)$ and where d is the last degree difference. Since in this case r does not involve t , the initial of $r(x, y)$ is $r(x, y)$ itself. Hence (5.2) becomes

$$(5.3) \quad \pm Ag(x, y)[f(x, y)]^k = [r(x, y)]^d.$$

It is clear from this relation that f divides r and that d (which is a multiple

of k) can exceed k only if f divides an initial of a polynomial preceding r . Hence we have

$$(5.4) \quad h(x, y)f(x, y) = r(x, y),$$

where f is the irreducible factor of the eliminant and h is a non-zero constant times the product of non-negative integral powers of initials of polynomials preceding r in the division sequence.

An important consequence is that *eliminating by the division process may introduce an extraneous factor $h(x, y)$, whereas eliminating by the eliminant never does.* An example follows.

$$\begin{aligned} x &= t^5 - t + 2, & y &= t^3 + t^2 + 2, \\ E(x, y) = f(x, y) &= 116 - 54x - 186y + 4x^2 + 67xy + 126y^2 + x^3 \\ &\quad - 5x^2y - 27xy^2 - 48y^3 + 5xy^3 + 10y^4 - y^5, \\ h(x, y) &= -(3 - y)^2. \end{aligned}$$

In the reducible case, the division process may yield the irreducible factor of $E(x, y)$, for example,

$$x = t^2 + t + 1, \quad y = t^2 + t + 2, \quad -x + y - 1 = r.$$

6. Polynomials in an unknown polynomial. Let us consider the reducibility of the parameter in the case of a single polynomial x . Since we may always write $x = u$, where u is of degree m , the parameter is always reducible, if the definition for reducibility is carried over directly from the case of two polynomials. Accordingly, we make the following definition: the parameter t in $x(t)$ is reducible if and only if x can be expressed as a polynomial in u , where u is a polynomial in t and $1 < \deg_t u < \deg_t x$. A parameter reducible for a pair of polynomials is then reducible for a polynomial x of the pair if and only if the irreducible factor of the eliminant is not linear in the other polynomial y of the pair. A single polynomial with parameter reducible in the sense just given is what Ritt [2] has called a *composite polynomial*.

In (1.1) let y be the polynomial with given complex coefficients. Let m be a fixed divisor of n satisfying $1 < m < n$. Let the coefficients of x be unknowns. The conditions that $E(x, y)$ be a perfect n/m power together with the inequation $a_0 \neq 0$ constitute an algebraic system which we shall denote by $S_m(y)$.

THEOREM 6.1. *The parameter in y is reducible if and only if $S_m(y)$ is consistent for some m satisfying*

$$m \mid \deg_i y, \quad 1 < m < \deg_i y.$$

Thus by discussing certain systems of algebraic equations it can be decided whether a reduced parameter exists and any reduced parameter can be determined.

DUKE UNIVERSITY

REFERENCES

1. O. Perron, *Algebra*, vol. 1, Berlin, 1932.
2. J. F. Ritt, "Prime and composite polynomials," *Transactions of the American Mathematical Society*, vol. 23 (1922), pp. 51-66.
3. J. M. Thomas, "Division sequences," *Duke Mathematical Journal*, vol. 13 (1946), pp. 459-469.

TAUBER'S THEOREM AND ABSOLUTE CONSTANTS.*

By PHILIP HARTMAN.

Let a_1, a_2, \dots be a sequence of real numbers. The theorem of Tauber [2] which states that an Abel summable series Σa_n is convergent when

$$a_1 + 2a_2 + \dots + na_n = o(n), \quad n \rightarrow \infty,$$

has recently been refined by Wintner [3] to a theorem involving an absolute constant as follows: There exists an absolute constant τ having the property that

$$(1) \quad \limsup_{r \rightarrow 1-0} \left| \sum_{m=1}^{\infty} a_m r^m - \sum_{m=1}^{-1/\log r} a_m \right| \leq \tau \limsup_{n \rightarrow \infty} |a_1 + 2a_2 + \dots + na_n|/n$$

for all sequences a_1, a_2, \dots . This implies the existence of another absolute constant τ^* such that

$$(2) \quad \limsup_{r \rightarrow 1-0} \left| \sum_{m=1}^{\infty} a_m r^m - \sum_{m=1}^{-1/\log r} a_m \right| \leq \tau^* \limsup_{n \rightarrow \infty} |na_n|$$

for all sequences a_1, a_2, \dots . The existence of this absolute constant implies the weaker theorem of Tauber that Abel summability and

$$na_n = o(1), \quad n \rightarrow \infty,$$

assure convergence. Hadwiger [1] had refined this last quoted theorem of Tauber in terms of an absolute constant ρ , which he defines in a somewhat different manner. However, it is clear from his proof for the existence of ρ that $\rho = \tau^*$. Hadwiger has obtained the following estimates for the best value of $\rho = \tau^*$,

$$0.4858 \dots \leq \tau^* \leq C + 2B = 1.01598 \dots,$$

where $C = 0.57721 \dots$ is the Euler-Mascheroni constant and

$$B = \int_1^{\infty} u^{-1} e^{-u} du = 0.21938 \dots,$$

so that

$$C + B = \int_0^1 u^{-1} (1 - e^{-u}) du = \sum_{n=1}^{\infty} (-1)^{n-1} / n \cdot n!.$$

Wintner has shown that

* Received November 23, 1946.

$$0.57721 \cdots = C \leq \tau^* \leq \tau$$

and his proof for the existence of τ gives the following estimate for the best value of τ :

$$1 \leq \tau \leq 3 + B = 3.21938 \cdots$$

The object of this paper is to obtain the best values for τ^* and τ . In particular, it will be shown that Hadwiger's upper bound for the best value of τ^* is actually the best value; so that (2) holds with

$$(3) \quad \tau^* = C + 2B = 1.01598 \cdots$$

for *all* sequences a_1, a_2, \cdots and the sign of equality holds for *some* sequences. On the other hand, the best value for τ is

$$(4) \quad \tau = C + 2B + 2/e = 1.75174 \cdots;$$

so that (4) satisfies (1) for *all* sequences a_1, a_2, \cdots and the sign of equality holds for *some* sequences.

The proofs of (3) and (4) will be based on several simple lemmas dealing with the geometric series and the power series for $-\log(1-r)$. In what follows, k, h, r are functions of n and all of the symbols $o(1)$ refer to $n \rightarrow \infty$.

LEMMA 1. If k_1, k_2, \cdots is an increasing sequence of integers such that

$$(5) \quad k = k_n = o(n),$$

and if r is a function of n satisfying

$$(6) \quad 0 < r < 1 \text{ and } -n \log r \rightarrow 1, \quad n \rightarrow \infty,$$

then, as $n \rightarrow \infty$,

$$(7) \quad \sum_{m=1}^k (1 - r^m)/m = o(1),$$

and

$$(8) \quad (1 - r) \sum_{m=1}^k r^m = o(1).$$

To prove (7), use will be made of the inequality

$$1 - r^m < m(1 - r), \quad (0 < r < 1),$$

a consequence of the mean value theorem of differential calculus. This inequality implies that the sum on the left of (7) is majorized by

$$\sum_{m=1}^k (1 - r) = k(1 - r),$$

which is $o(1)$ in virtue of (5), since (6) is equivalent to

$$(9) \quad 1 - r = 1/n + o(1/n).$$

To prove (8), it is sufficient to note that

$$(1 - r) \sum_{m=0}^k r^m = 1 - r^{k+1} < (k+1)(1 - r).$$

Hence (8) follows from (9) and (5).

LEMMA 2. *If r is a function of n satisfying (6), then*

$$(10) \quad \sum_{m=1}^n (1 - r^m)/m = C + B + o(1),$$

and

$$(11) \quad (1 - r) \sum_{m=1}^n r^m = 1 - 1/e + o(1).$$

The sum on the left of (11) is, up to a factor r ,

$$1 - r^n = 1 - (1 - 1/n + o(1/n))^n$$

by (9). Hence (11) follows at once.

The sum on the left of (10) can be written as

$$\sum_{m=1}^n 1/m - \sum_{m=1}^{\infty} r^m/m + \sum_{m=n+1}^{\infty} r^m/m.$$

Since the first term of this expression is $\log n + C + o(1)$, and the second is $\log(1 - r) = -\log n + o(1)$, it is sufficient to prove the following:

LEMMA 3. *If r is a function of n satisfying (6), then, as $n \rightarrow \infty$,*

$$(12) \quad \sum_{m=n+1}^{\infty} r^m/m = B + o(1),$$

and

$$(13) \quad (1 - r) \sum_{m=n+1}^{\infty} r^m = 1/e + o(1).$$

First, (13) is an immediate consequence of (11) since

$$(1 - r) \sum_{m=0}^{\infty} r^m = 1.$$

Next, the sum on the left of (12) can be written as the integral

$$\int_0^r t^n (1 - t)^{-1} dt = \int_{-n \log r}^{\infty} n^{-1} (e^{u/n} - 1)^{-1} e^{-u} du,$$

where the last integral is obtained by an obvious change of variable. Since $n(e^{u/n} - 1) \rightarrow u$, $n \rightarrow \infty$, uniformly on any bounded u -interval, since the integrand of the last integral is majorized by e^{-u}/u for all values of $u > 0$ and, finally, since (6) holds, the relation (12) is a consequence of the last formula and the definition of B .

LEMMA 4. *If h_1, h_2, \dots is a sequence of increasing positive integers such that*

$$(14) \quad n = o(h), \text{ where } h = h_n,$$

and if r is a function of n satisfying (6), then, as $n \rightarrow \infty$,

$$(15) \quad \sum_{m=h}^{\infty} r^m/m = o(1),$$

and

$$(16) \quad (1-r) \sum_{m=h}^{\infty} r^m = o(1).$$

The sum occurring in (15) can be written as an integral and appraised as follows

$$\int_0^r t^h (1-t)^{-1} dt < \int_{-h \log r}^{\infty} u^{-1} e^{-u} du.$$

Since (6) and (14) imply that

$$(17) \quad -h \log r \rightarrow \infty, \quad n \rightarrow \infty,$$

(15) follows. As to the statement (16), it is sufficient to observe that the sum on the left of (16) is $r^h = o(1)$ by (17).

The proof of (3) will now be given. It will first be shown that

$$(18) \quad \tau^* \geq C + 2B.$$

The difference

$$(19) \quad \sum_{m=1}^{\infty} a_m r^m - \sum_{m=1}^n a_m$$

can be written in the form

$$\sum_{m=1}^k m a_m (r^m - 1)/m + \sum_{m=k+1}^n m a_m (r^m - 1)/m + \sum_{m=n+1}^h m a_m r^m/m + \sum_{m=h+1}^{\infty} m a_m r^m/m,$$

where k and h are integers such that $1 < k < n < h$. Let $0 = k_0, k_1, k_2, \dots$ be an increasing sequence of integers such that

$$(20) \quad k_{n-1} = o(k_n), \quad n \rightarrow \infty,$$

and let the sequence of numbers a_1, a_2, \dots be defined by placing

$$(21) \quad ma_m = (-1)^j \text{ if } k_j < m \leq k_{j+1} \quad (j = 0, 1, \dots).$$

The corresponding difference (19), with $n = k_j$ and r a function of n satisfying (6), can be written in the form following (19), with $k = k_{j-1}$ and $h = k_{j+1}$. Then (21), (20) and Lemma 1 imply that

$$(22) \quad \sum_{m=1}^k ma_m(r^m - 1)/m = o(1);$$

while (21) and Lemma 1 give

$$\begin{aligned} \sum_{m=k+1}^n ma_m(r^m - 1)/m &= (-1)^j \sum_{m=k+1}^n (1 - r^m)/r^m \\ &= (-1)^j \sum_{m=1}^n (1 - r^m)/m + o(1). \end{aligned}$$

Hence, by Lemma 2,

$$(23) \quad \sum_{m=k+1}^n ma_m(r^m - 1)/m = (-1)^j(C + B) + o(1).$$

Similarly, Lemma 4 implies that

$$(24) \quad \sum_{m=h+1}^{\infty} ma_m r^m/m = o(1),$$

and that

$$\sum_{m=n+1}^h ma_m r^m/m = (-1)^j \sum_{j=n+1}^{\infty} r^m/m + o(1).$$

Hence, by Lemma 3,

$$(25) \quad \sum_{m=n+1}^h ma_m r^m/m = (-1)^j B + o(1).$$

The relations (22), (23), (24) and (25) show that if $n = k_j$ and if r is a function of n satisfying (6), then the difference (19) is

$$(-1)^j(C + 2B) + o(1), \quad n \rightarrow \infty.$$

Hence, (18) follows from (2) and (21).

Accordingly, to complete the proof of (3), it is sufficient to show that

$$\tau^* \leq C + 2B.$$

But this inequality follows from Hadwiger's upper estimate for τ^* .

In order to verify (4), first

$$(26) \quad \tau \leq C + 2B + 2/e$$

will be proved.

Let a_1, a_2, \dots be an arbitrary sequence of real numbers, and put

$$(27) \quad t_n = \sum_{m=1}^n m a_m / n, \quad n = 1, 2, \dots$$

It can be supposed that

$$(28) \quad L = \limsup_{n \rightarrow \infty} |t_n| < \infty,$$

since otherwise (1) is trivial. Let

$$(29) \quad L_n = \limsup_{m > n} |t_m|;$$

so that

$$(30) \quad L_n \rightarrow L, \quad n \rightarrow \infty.$$

By Abel's identity,

$$\sum_{m=1}^{\infty} a_m r^m = \sum_{m=1}^{\infty} m a_m r^m / m = \sum_{m=1}^{\infty} m t_m (r^m / m - r^{m+1} / (m+1)).$$

This can be simplified to

$$\sum_{m=1}^{\infty} a_m r^m = \sum_{m=1}^{\infty} t_m (r^m (1-r) + r^{m+1} / (m+1)).$$

Also, by (27) and Abel's identity,

$$\sum_{m=1}^n a_m = \sum_{m=1}^n m a_m / m = \sum_{m=1}^n t_m / (m+1) + n t_n / (n+1).$$

Hence, the difference (19) can be written as the sum of the following five expressions:

$$(31_1) \quad \sum_{m=1}^k t_m (r^m (1-r) + (r^{m+1} - 1) / (m+1)),$$

$$(31_2) \quad \sum_{m=k+1}^n t_m (r^m (1-r) + (r^{m+1} - 1) / (m+1)),$$

$$(31_3) \quad (1-r) \sum_{m=n+1}^{\infty} t_m r^m,$$

$$(31_4) \quad \sum_{m=n+1}^{\infty} t_m r^{m+1} / (m+1),$$

$$(31_5) \quad -n t_n / (n+1),$$

where k is an integer such that $1 < k < n$.

Let $k = k_n$ be a function of n satisfying (5) and

$$(32) \quad k_n \rightarrow \infty, \quad n \rightarrow \infty,$$

and let r be a function of n satisfying (6). Then (28) and Lemma 1 imply that the sum (31₁) is $o(1)$ as $n \rightarrow \infty$.

To treat the sum in (31₁), it is important to notice that, since $0 < r < 1$,
 $|r^m(1-r) + (r^{m+1}-1)/(m+1)| = (1-r^{m+1})/(m+1) - r^m(1-r)$,
 in view of

$$\begin{aligned} & r^m(1-r) + (r^{m+1}-1)/(m+1) \\ &= (1-r)((m+1)r^m - \sum_{j=0}^m r^j)/(m+1) < 0. \end{aligned}$$

Hence, the sum (31₂) is majorized by

$$L_k \left(\sum_{m=k+1}^n (1-r^m)/m - (1-r) \sum_{n=k+1}^n r^m \right).$$

In virtue of (5), (6) and Lemma 1, the factor of L_k is

$$\sum_{m=1}^n (1-r^m)/m - (1-r) \sum_{m=1}^n r^m + o(1),$$

which, by Lemma 2, is

$$(33_2) \quad C + B - 1 + 1/e + o(1).$$

In virtue of (29) and Lemma 3, the absolute values of the sums (31₃) and (31₄) can be appraised by

$$(33_3) \quad L_n(1-r) \sum_{m=n+1}^{\infty} r^m = L_n(1/e + o(1))$$

and

$$(33_4) \quad L_n \sum_{m=n+1}^{\infty} r^m/m = L_n(B + o(1)),$$

respectively.

Finally, the absolute value of (31₅) does not exceed

$$(33_5) \quad L_n(1 + o(1)).$$

Collecting the results (33₂)-(33₅), it is seen that the difference (19) is majorized by

$$L_k(C + B - 1 + 1/e) + L_n(1/e + B + 1) + o(1).$$

In view of (29), (30) and (32), it follows that (1) holds for all sequences a_1, a_2, \dots when τ is defined by (4). This proves (26).

In order to complete the proof of (4), it remains to be shown that

$$(34) \quad \tau \geq C + 2B + 2/e.$$

To this end let $0 = k_0, k_1, k_2, \dots$ be an increasing sequence of integers satisfying (20). Define a sequence a_1, a_2, \dots in terms of the numbers t_1, t_2, \dots by placing

$$(35) \quad t_m = (-1)^j \text{ if } k_j < m \leq k_{j+1} \quad (j = 0, 1, 2, \dots).$$

The detailed computations may be omitted. In fact, it is clear from the above construction that the difference (19), with $n = k_j$ and with r as a function of n satisfying (6), is

$$(-1)^j(C + 2B + 2/e) + o(1).$$

This implies (34) in virtue of (1) and (35).

The proof for the existence of τ given by Wintner treated the Laplace transform analogue of (1) rather than (1) itself. Actually, the above considerations are easily transcribed to this more general situation. Let $\alpha(x)$ be defined for all $x \geq 1$ and of bounded variation on any bounded x -interval. Define $\beta(x)$ by

$$\beta(x) = \int_1^x t d\alpha(t) \text{ if } x \geq 1.$$

Then

$$\limsup_{s \rightarrow +0} \left| \int_1^\infty e^{-st} d\alpha(t) - \int_1^{1/s} d\alpha(t) \right| \leq \tau \limsup_{x \rightarrow \infty} |\beta(x)/x|$$

holds if τ is the number (4).

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES

- [1] H. Hadwiger, "Ueber ein Distanztheorem bei der A -Limitierung," *Commentarii Mathematici Helvetici*, vol. 16 (1943-44), pp. 209-214.
- [2] A. Tauber, "Ein Satz aus der Theorie unendlichen Reihen," *Monatshefte für Mathematik und Physik*, vol. 8 (1897), pp. 273-277.
- [3] A. Wintner, "On Tauber's theorem."

PROJECTIVE THEORY OF SURFACES AND CONJUGATE NETS IN FOUR-DIMENSIONAL SPACE.*

By CHUAN-CHIH HSIUNG.

Introduction. It is known that every non-developable surface immersed in a linear space S_4 of four dimensions has on it either a conjugate net, or else a unique one-parameter family of asymptotic curves. The purpose of this paper is to establish a purely geometric theory of the projective differential geometry of the surfaces sustaining a conjugate net.¹

In **1** a completely integrable system of linear homogeneous partial differential equations, together with its canonical form by a geometric determination, is introduced which defines a conjugate net in space S_4 except for a projective transformation.

In **2** we study the effect on the differential equations of **1** of a group of transformations which leave invariant the parametric conjugate net N_x on an integral surface S of these equations. Some invariants and covariants of this system under the group of transformations are also obtained and listed.

In **3** we calculate for the surface S local power series expansions, which express two local non-homogeneous projective coordinates of a point on the surface S as two power series in the other two coordinates and represent the surface S in the neighborhood of an ordinary point on it.

4 is devoted to the derivation of a certain cubic hypercone associated with a point of the surface S .

In **5**, making use of the hyperquadrics having third order contact with the surface S at a point x , we define a *reciprocal correspondence* between certain lines in the tangent plane of the surface S at the point x and certain planes passing through the point x , and call these lines *the canonical lines*

* Received December 10, 1946; Presented to the American Mathematical Society, August 22, 1946.

¹ A projective theory of conjugate nets in ordinary three-dimensional space has been established in a similar way. See E. P. Lane, *A Treatise on Projective Differential Geometry*, The University of Chicago Press, 1942, Chap. VIII.

and these planes *the canonical planes* of the conjugate net N_x at the point x . All canonical lines pass through a point called *the canonical point*, and all canonical planes lie on a quadric hypercone passing through the canonical point called *the canonical quadric hypercone*. The tangent hyperplane of the canonical quadric hypercone at the canonical point is called *the canonical hyperplane*.

6 contains the local power series expansions for the u -, v -curves of the parametric conjugate net N_x on the surface S . These expansions express three local non-homogeneous projective coordinates of a point on each curve as power series in the other coordinate, and represent the curve in the neighborhood of an ordinary point on it. Among the hyperquadrics having third order contact with the surface S at a point x we may determine a pencil of hyperquadrics such that they have fourth order contact with the v -curve and fifth order contact with the u -curve at the point x . The equations of this pencil of hyperquadrics are included at the end of the section.

It is known that as u, v vary the Laplace transformed points x_{-1}, x_1 of an ordinary point x of the surface S with respect to the v -, u -curves of the conjugate net N_x generate two surfaces S_{-1}, S_1 , on which the parametric curves also form two conjugate nets N_{-1}, N_1 . In the last section geometric characterizations of some special classes of conjugate nets are obtained by studying the pencil of hyperquadrics which have third order contact with the surface $S_{-1}(S_1)$ at the point $x_{-1}(x_1)$ and first order contact with the surface $S_1(S_{-1})$ at the point $x_1(x_{-1})$.

1. Differential equations and integrability conditions. Let us consider in a linear space S_4 of four dimensions a non-developable surface S sustaining a conjugate net N_x with the parametric vector equation, referred to the net N_x ,

$$(1.1) \quad x = x(u, v).$$

The osculating space S_3 of the parametric curve $u(v)$ and the osculating plane of the parametric curve $v(u)$ at an ordinary point x of the surface S intersect in a line $l_1(l_2)$. Let us select on the lines l_1 and l_2 respectively two points y and z , distinct from the point x , and suppose that the coordinates y and z of the points y and z are functions of u, v . Then it can be shown that the coordinates x, y and z of corresponding points x, y and z satisfy a system of linear homogeneous partial differential equations of the form

$$\begin{aligned}
 (1.2) \quad & x_{uu} = px + \alpha x_u + Lz, \\
 & x_{uv} = cx + ax_u + bx_v, \\
 & x_{vv} = qx + \delta x_v + Ny, \\
 & x_{uuu} = rx + mx_u + \rho y + \beta z, \\
 & x_{vvv} = sx + nx_v + \gamma y + \sigma z \quad (\rho\sigma LN \neq 0),
 \end{aligned}$$

in which subscripts indicate partial differentiation and the coefficients are scalar functions of u, v . The second of these equations is merely the Laplace equation for the parametric conjugate net N_x .

It is easy to express the derivatives y_u, y_v as linear combinations of x, x_u, x_v, y, z by using equations (1.2) and

$$(x_{uv})_v = (x_{vv})_u, \quad (x_{vv})_v = x_{vvv};$$

the result is

$$\begin{aligned}
 (1.3) \quad & y_u = ex + Rx_u + Ex_v + Ay, \\
 & y_v = gx + Px_v + By + (\sigma/N)z,
 \end{aligned}$$

in which the coefficients are defined by the following equations:

$$\begin{aligned}
 (1.4) \quad & eN = c_v + ac + bq - c\delta - q_u, & gN = s - q\delta - q_v, \\
 & RN = a_v + a^2 - a\delta - q, & PN = n - q - \delta_v - \delta^2, \\
 & EN = b_v + ab + c - \delta_u, & BN = \gamma - \delta N - N_v, \\
 & A = b - (\log N)_u.
 \end{aligned}$$

Analogous expressions for z_u, z_v can be written by making the substitution

$$(1.5) \quad \begin{pmatrix} u & x_1 & y & a & c & m & p & r & \alpha & \beta & \rho & L & e & g & A & B & E & P & R \\ v & x_1 & z & b & c & n & q & s & \delta & \gamma & \sigma & N & f & h & D & C & F & Q & S \end{pmatrix}.$$

We now proceed to choose for the points y and z two particular covariant points on the lines l_1 and l_2 respectively. To this end we observe that the point X defined by

$$X = y + kx$$

where k is a scalar function of u, v , is on the line l_1 . When the point x varies along a curve C_λ of the family represented by the differential equation

$$(1.6) \quad dv - \lambda du = 0,$$

λ being a function of u, v , on the surface S , the point X generates a curve C_X whose tangent at X is determined by X and the point X' given by

$$X' = y_u + y_v\lambda + k(x_u + x_v\lambda) + k'x \quad (X' = dX/du, \dots).$$

Expressing X' as a linear combination of x, x_u, x_v, y, z by means of equations (1.3), and equating to zero the coefficients of x_u, x_v therein, we obtain two conditions on the functions k and λ which are necessary and sufficient that the tangent to the curve C_X at the point X lies in the plane $l_1 l_2$, namely,

$$k + R = 0, \quad E + (P + k)\lambda = 0.$$

Similarly, we can also determine a unique point on the line l_2 and a unique curve of the family (1.6) connected with the point. If we choose these two points respectively for the points y and z , then

$$(1.7) \quad R = 0, \quad S = 0,$$

and the differential equations of the two curves become

$$(1.8) \quad Pdv + Edu = 0, \quad Fdv + Qdu = 0.$$

Hereafter it will be supposed that the differential equations (1.2) are in the *canonical form* for which the conditions (1.7) are satisfied.

The integrability conditions of equations (1.2) are found by the usual method from the equations

$$(y_u)_v = (y_v)_u, \quad (z_u)_v = (z_v)_u$$

and the fact that the points x, x_u, x_v, y, z are linearly independent. These conditions are

$$(1.9) \quad \begin{aligned} e_v + gA + qE &= g_u + eB + cP + h\sigma/N, & g + aP + \sigma Q/N &= 0, \\ e + E_v + \delta E + AP &= P_u + bP + BE, & A_v + EN &= B_u + \rho\sigma/LN \\ \sigma_u &= \sigma(b - c), \end{aligned}$$

and the analogous ones obtainable therefrom by the substitution (1.5).

Making use of equations (1.4), the fourth of equations (1.9) and the substitution (1.5) we obtain the equation

$$(1.10) \quad (2b + \beta/L)_v = (2a + \gamma/N)_u.$$

It follows that there exists a function θ of u, v which is defined, except for an arbitrary additive constant, as a solution of the differential equations

$$(1.11) \quad \theta_u = 2b + \beta/L - (\log LN)_u, \quad \theta_v = 2a + \gamma/N - (\log LN)_v.$$

Accordingly, the following formula is valid:

$$(1.12) \quad (x, x_u, x_v, y, z) = e^\theta,$$

where a determinant is indicated by writing only a typical row within parentheses.

2. Transformations and invariants. Let us consider the group of transformations on the coordinates x, y, z and the parameters u, v :

$$(2.1) \quad x = \lambda \bar{x}, \quad y = \mu \bar{y}, \quad z = \nu \bar{z} \quad (\lambda \mu \nu \neq 0),$$

$$(2.2) \quad \bar{u} = U(u), \quad \bar{v} = V(v) \quad (U'V' \neq 0),$$

where λ, μ, ν are scalar functions of u, v and the accent denotes differentiation with respect to the appropriate variable.

The effect of the transformation (2.1) on the system of equations (1.2) is to produce another system of equations of the same form whose coefficients, indicated by dashes, are given by the following formulas and the analogous ones:

$$(2.3) \quad \begin{aligned} \bar{p} &= (1/\lambda)(p\lambda + \alpha\lambda_u - \lambda_{uu}), & \bar{\alpha} &= \alpha - 2(\lambda_u/\lambda), & \bar{L} &= (v/\lambda)L, \\ \bar{c} &= (1/\lambda)(c\lambda + a\lambda_u + b\lambda_v - \lambda_{uv}), & \bar{a} &= a - \lambda_v/\lambda, \\ \bar{r} &= r + (1/\lambda)(m - 3p)\lambda_u + 3(\lambda_u/\lambda^2)(\lambda_{uu} - \alpha\lambda_u) - \lambda_{uuu}/\lambda, \\ \bar{m} &= m - (3/\lambda)(\alpha\lambda_u + \lambda_{uu}) + 6(\lambda_u^2/\lambda^2), & \bar{\rho} &= (\mu/\lambda)\rho, \\ \bar{\beta} &= (v/\lambda^2)(\beta\lambda - 3L\lambda_u). \end{aligned}$$

The effect of the transformation (2.2) on the system of equations (1.2) is to produce another system of equations of the same form whose coefficients, indicated by stars, are given by the following formulas and the analogous ones:

$$(2.4) \quad \begin{aligned} p^* &= (1/U'^2)p, & \alpha^* &= (1/U')(\alpha - U''/U'), & L^* &= (1/U'^2)L, \\ c^* &= (1/U'V')c, & a^* &= (1/V')a, \\ r^* &= (1/U'^3)[r - 3(U''/U')p], \\ m^* &= (1/U'^2)[m - (1/U')(3\alpha U'' + U''') + 3(U''/U')^2], \\ \rho^* &= (1/U'^3)\rho, & \beta^* &= (1/U'^3)[\beta - 3(U''/U')L]. \end{aligned}$$

From equations (2.3), (2.4) and the substitution (1.5) we may obtain, after some calculation, the following functions which are absolute invariants, under the transformation (2.1), and relative invariants, under the transformation (2.2), of the system of equations (1.2):

$$(2.5) \quad \begin{aligned} H &= c + ab - a_u, & \mathfrak{H} &= 3(b - \alpha) + \beta/L, \\ I &= L/\sigma, & \mathfrak{M} &= 4b - 2\alpha + \phi_u, & \mathfrak{S} &= E/N, \\ \tilde{\gamma} &= r + bm - b^3 - 3bb_u - b_{uu}, & \mathfrak{R} &= 3(2b - \alpha) + [\log(\rho/N)]_u, \\ \mathfrak{B} &= -3m + 3(6b^2 - 10b\alpha + 5\alpha^2) + 6b_u + (\beta/L)(8b - 7\alpha) \\ &\quad + \beta^2/L^2 + (\beta/L)_u, \end{aligned}$$

where $\phi = \log I$.

The effect of the transformation (2.2) on each of these invariants is given, with self-explanatory notation, by the following formulas:

$$(2.6) \quad \begin{aligned} H^* &= (1/U'V')H, & \mathfrak{H}^* &= (1/U')\mathfrak{H}, \\ I^* &= (V'^3/U'^2)I, & \mathfrak{M}^* &= (1/U')\mathfrak{M}, & \mathfrak{S}^* &= (1/U'V')\mathfrak{S}, \\ \mathfrak{F}^* &= (1/U'^3)\mathfrak{F}, & \mathfrak{N}^* &= (1/U')\mathfrak{N}, & \mathfrak{B}^* &= (1/U'^2)\mathfrak{B}. \end{aligned}$$

Analogous invariants can be written by using the substitution (1.5) and

$$(2.7) \quad \begin{pmatrix} H & I & \mathfrak{N} & \mathfrak{B} & \mathfrak{F} & \mathfrak{S} & \phi & \mathfrak{M} & \mathfrak{S} \\ K & J & \mathfrak{D} & \mathfrak{C} & \mathfrak{G} & \mathfrak{R} & \psi & \mathfrak{N} & \mathfrak{L} \end{pmatrix}.$$

Two of the most important covariants of the system of equations (1.2) under the total transformation (2.1), (2.2) are the functions x_1, x_{-1} defined by the formulas

$$(2.8) \quad x_1 = x_v - ax, \quad x_{-1} = x_u - bx.$$

The points x_1, x_{-1} are called, respectively, the *first and minus-first Laplace transformed points* of the point x with respect to the parametric conjugate net N_x on the surface S , or the *ray-points* of the u, v -curves at the point x .

3. Power series expansions for a surface. Let us consider an ordinary point x with curvilinear coordinates u, v on an integral surface S of equations (1.2) with the conditions (1.7). The coordinates X , where

$$X = x(u + \Delta u, v + \Delta v),$$

of any point X near the point x on the surface S are given by the Taylor's series

$$(3.1) \quad \begin{aligned} X &= x + x_u \Delta u + x_v \Delta v \\ &\quad + \frac{1}{2}(x_{uu} \Delta u^2 + 2x_{uv} \Delta u \Delta v + x_{vv} \Delta v^2) + \cdots, \end{aligned}$$

in which the increments Δu and Δv correspond to displacement on the surface S from the point x to the point X . By means of equations (1.2), (1.3), the equations obtained therefrom by differentiation and the substitution (1.5), it is possible to express every derivative of x uniquely as a linear combination of x, x_u, x_v, y, z .

If the points x, x_{-1}, x_1, y, z are used as the vertices of the pyramid of reference, with unit point suitably chosen, then any point given by an expression of the form

$$(3.2) \quad \xi_1 x + \xi_2 x_{-1} + \xi_3 x_1 + \xi_4 y + \xi_5 z$$

has local coordinates proportional to ξ_1, \dots, ξ_5 . Replacing the derivatives x_u, x_v in the foregoing formulas by their equivalents as given by equations (2.8), we are able easily to express derivatives of x of any required order as linear combinations of x, x_{-1}, x_1, y, z . Then substituting these expressions for the derivatives of x in the Taylor's series (3.1) and making use of equations (1.4), (2.5), the substitution (2.7) and the definitions

$$(3.3) \quad \xi = \xi_2/\xi_1, \quad \eta = \xi_3/\xi_1, \quad \zeta = \xi_4/\xi_1, \quad \tau = \xi_5/\xi_1,$$

we may obtain the power series expansions for the non-homogeneous local coordinates ζ, τ of the point X in terms of the other two coordinates ξ, η :

$$(3.4) \quad \begin{aligned} \zeta &= \frac{1}{2}N\eta^2 + \frac{1}{6}\rho\xi^3 + \frac{1}{6}N\mathfrak{R}\eta^3 + (\rho/24)(\mathfrak{M} + \mathfrak{S})\xi^4 \\ &\quad + \frac{1}{6}NK\xi\eta^3 + (1/24)N\mathfrak{C}\eta^4 + \dots, \\ \tau &= \frac{1}{2}L\xi^2 + \frac{1}{6}L\mathfrak{S}\xi^3 + \frac{1}{6}\sigma\eta^3 + (1/24)L\mathfrak{B}\xi^4 \\ &\quad + \frac{1}{6}LII\xi^3\eta + (\sigma/24)(\mathfrak{D} + \mathfrak{R})\eta^4 + \dots \end{aligned}$$

The first of equations (3.4) and $\tau = 0$ demonstrate that the projection of the surface S from the point z into the space S_3 $xx_{-1}x_1y$ has coincident asymptotic tangents at the point x . A similar statement holds for the second of equations (3.4) and $\zeta = 0$.

4. A certain cubic hypercone. In this section we shall derive a certain cubic hypercone by considering a general curve C_λ represented by the differential equation (1.6). The first three derivatives of x with respect to the independent variable u along the curve C_λ are found to be

$$(4.1) \quad \begin{aligned} x' &= x_u + x_v\lambda & (x' = dx/du, \dots), \\ x'' &= G + x_v\lambda', \\ x''' &= H + 3(x_{uv} + x_{vv}\lambda)\lambda' + x_v\lambda'', \end{aligned}$$

where G, H are defined by placing

$$(4.2) \quad \begin{aligned} G &= x_{uu} + 2x_{uv}\lambda + x_{vv}\lambda^2, \\ H &= x_{uuu} + 3x_{uuv}\lambda + 3x_{uvv}\lambda^2 + x_{vvv}\lambda^3. \end{aligned}$$

The space $S(2, 1)$ of the surface S in the direction of the curve C_λ at the point x , which is defined as the space of least dimensionality containing the osculating planes at the point x of all curves lying on the surface S and tangent to the curve C_λ at the point x , is a hyperplane determined by the points

$$(4.3) \quad x, x_u, x_v, G.$$

The differential equation of the two-parameter family of the quasi-asymptotic curves of Bompiani² at the point x of the surface S now takes the form

$$(4.4) \quad 3LN\lambda\lambda' + \rho L + N(3bL - \beta)\lambda^2 + L(\gamma - 3aN)\lambda^3 - \sigma N\lambda^5 = 0.$$

Since the equation of the space $S(2, 1)$ in local coordinates is

$$(4.5) \quad L\xi_4 - N\lambda^2\xi_5 = 0,$$

we may easily verify that *every hyperplane through the tangent plane at the point x of the surface S is the osculating space S_3 of two of these quasi-asymptotic curves at the point x , whose directions separate the u -, v -tangents at the point x harmonically.*³

The tangent line of the curve C_λ and the plane of intersection of the two osculating spaces S_3 of the u -, v -curves at the point x determine a hyperplane with the equation

$$(4.6) \quad \xi_3 - \lambda\xi_2 = 0.$$

As the direction (1.6) varies in the tangent plane of the surface S at the point x , the locus of the plane of intersection of the two hyperplanes (4.5), (4.6) is a cubic hypercone with vertex at the point x , whose equation is found to be

$$(4.7) \quad L\xi_2^2\xi_4 - N\xi_3^2\xi_5 = 0.$$

This cubic hypercone has $xx_{-1}z$, xx_1y for its double planes and passes through the tangent plane $xx_{-1}x_1$.

5. Canonical configurations for a surface. There is a four-parameter family of hyperquadrics having third order contact with the surface S at a point x . The equation of a general one of these hyperquadrics is obtained by writing the equation of the most general non-singular hyperquadric and demanding that the series (3.4) satisfy this equation identically in ξ , η as far as the terms of the third degree. The result can be written in the form

$$(5.1) \quad \frac{1}{2}La_{45}\xi^2 + (1/3L)(\rho a_{35} + L\mathfrak{S}a_{45})\xi\tau + \frac{1}{2}Na_{35}\eta^2 \\ + (1/3N)(N\mathfrak{R}a_{35} + \sigma a_{45})\eta\zeta + a_{33}\zeta^2 + a_{34}\zeta\tau \\ + a_{44}\tau^2 - a_{35}\xi - a_{45}\tau = 0,$$

² E. Bompiani, "Sopra alcune estensioni dei teoremi di Meusnier di Eulero," *Atti della Reale Accademia delle Scienze di Torino*, vol. 48 (1912), p. 404; or see E. P. Lane, *loc. cit.*¹, p. 414.

³ E. Bompiani, *loc. cit.*²

where the coefficients a_{ij} are parameters. The polar hyperplanes of the points x_{-1} , x_1 with respect to any hyperquadric of the family (5.1) have respectively the equations

$$(5.2) \quad La_{45}\xi + (1/3L)(\rho a_{35} + L\mathfrak{S}a_{45})\tau = 0,$$

$$(5.3) \quad Na_{35}\eta + (1/3N)(N\mathfrak{R}a_{35} + \sigma a_{45})\xi = 0.$$

Let us now consider in the tangent plane $xx_{-1}x_1$ two directions which pass through the point x and separate the u -, v -tangents harmonically. The equations of these directions may be written in the form

$$(5.4) \quad \eta^2 - \lambda^2\xi^2 = 0, \quad \xi = 0, \quad \tau = 0,$$

where λ is arbitrary. If the hyperquadric (5.1) cuts the tangent plane $xx_{-1}x_1$ in the directions (5.4), the polar hyperplanes (5.2), (5.3) become respectively

$$(5.5) \quad N\lambda^2\xi + (1/3L)(N\mathfrak{S}\lambda^2 - \rho)\tau = 0.$$

$$(5.6) \quad L\eta + (1/3N)(L\mathfrak{R} - \sigma\lambda^2)\xi = 0.$$

As the tangents (5.4) vary about the point x the plane of intersection of the two hyperplanes (5.5), (5.6) generates a quadric hypercone with vertex at the point x , whose equation is

$$(5.7) \quad [\xi + (\mathfrak{S}/3L)\tau][\eta + (\mathfrak{R}/3N)\xi] = (\rho\sigma/9L^2N^2)\xi\tau.$$

It is evident that the polar plane of the tangent plane $xx_{-1}x_1$ with respect to the hypercone (5.7) is

$$(5.8) \quad \xi + (\mathfrak{S}/3L)\tau = 0, \quad \eta + (\mathfrak{R}/3N)\xi = 0.$$

If the line yz lies on the hyperquadrics (5.1), then $a_{33} = a_{34} = a_{44} = 0$ and therefore we obtain a pencil of hyperquadrics:

$$(5.9) \quad a_{35}[\xi - (\rho/3L)\xi\tau - (N/2)\eta^2 - (\mathfrak{R}/3)\eta\xi] \\ + a_{45}[\tau - (\mathfrak{S}/3)\xi\tau - (L/2)\xi^2 - (\sigma/3N)\eta\xi] = 0.$$

In this pencil there is a unique pair, one passing through the points x_{-1} , z and the other through the points x_1 , y , whose equations are respectively

$$(5.10) \quad \xi - (\rho/3L)\xi\tau - (N/2)\eta^2 - (\mathfrak{R}/3)\eta\xi = 0,$$

$$(5.11) \quad \tau - (\mathfrak{S}/3)\xi\tau - (L/2)\xi^2 - (\sigma/3N)\eta\xi = 0.$$

The polar hyperplanes p_2 , p_1 of the points x_1 , x_{-1} with respect to the hyperquadrics (5.10), (5.11), respectively, are given by equations (5.8). More-

over, the polar hyperplanes of the points y, z with respect to the hyperquadrics (5.10), (5.11), are, respectively,

$$(5.12) \quad 3 - \mathfrak{R}\eta = 0, \quad 3 - \mathfrak{S}\xi = 0,$$

which intersect the v -, u -tangents respectively in the points

$$(5.13) \quad \mathfrak{R}x + 3x_1, \quad \mathfrak{S}x + 3x_{-1}.$$

Let P_2, P_1 be the harmonic conjugate points of the two points (5.13) with respect to the points x, x_1 and x, x_{-1} respectively, then on the lines xx_{-1}, xx_1 we may determine two points Q_1, Q_2 such that the cross ratios

$$(5.14) \quad (xx_{-1}, P_1Q_1) = 3k, \quad (xx_1, P_2Q_2) = 3k,$$

where k is an arbitrary constant independent of u, v . The equations of the line Q_1Q_2 are of the form

$$(5.15) \quad 1 + k\mathfrak{S}\xi + k\mathfrak{R}\eta = 0, \quad \xi = 0, \quad \tau = 0.$$

In like manner, we can determine two hyperplanes q_1, q_2 such that the cross ratios

$$(5.16) \quad (xx_{-1}x_1y, xx_1yz, p_1, q_1) = 3k, \quad (xx_{-1}x_1z, xx_{-1}yz, p_2, q_2) = 3k.$$

The plane of intersection of the two hyperplanes q_1, q_2 is given by the equations

$$(5.17) \quad L\xi + k\mathfrak{S}\tau = 0, \quad N\eta + k\mathfrak{R}\xi = 0.$$

The line Q_1Q_2 and the plane q_1q_2 will be said to be *reciprocal* with respect to the net N_x . The line Q_1Q_2 will be called a *canonical line* and the plane q_1q_2 a *canonical plane* of the net N_x at the point x . All canonical lines pass through the point

$$(5.18) \quad (0, \mathfrak{R}, -\mathfrak{S}, 0, 0),$$

which we shall call *the canonical point* at the point x of the net N_x . All canonical planes lie on the quadric hypercone

$$(5.19) \quad L\mathfrak{R}\xi\xi - N\mathfrak{S}\eta\tau = 0,$$

which we shall call *the canonical quadric hypercone* at the point x of the net N_x . The tangent hyperplane of the canonical quadric hypercone at the canonical point will be called *the canonical hyperplane* at the point x of the net N_x ; its equation is

$$(5.20) \quad L\mathfrak{R}^2\xi + N\mathfrak{S}^2\tau = 0.$$

If the plane of intersection of the two polar hyperplanes (5.2), (5.3)

is a general canonical plane (5.17), then the conjugate net N_x is restricted by the condition

$$(5.21) \quad (3k-1)^2 IJ\mathfrak{S}\mathfrak{R} = 1.$$

Moreover, we may determine two pairs of tangents through the point x such that the plane of intersection of the two polar hyperplanes (5.5), (5.6) is a general canonical plane (5.17). The equations of these tangents are easily found to be

$$(5.22) \quad (1-3k)N\mathfrak{S}\eta^2 - \rho\xi^2 = 0, \quad \xi = 0, \quad \tau = 0,$$

$$(5.23) \quad \sigma\eta^2 - (1-3k)L\mathfrak{R}\xi^2 = 0, \quad \xi = 0, \quad \tau = 0.$$

6. Curves of a conjugate net. In this section we shall study the u -, v -curves of the parametric conjugate net N_x on an integral surface S of equations (1.2). For this purpose we first consider a point X near a point x and on the u -curve through x , whose coordinates are given by the Taylor's series in the increment Δu corresponding to displacement along the u -curve from the point x to the point X . It is easy to express the non-homogeneous local coordinates η, ξ, τ of the point X as power series in the other coordinate ξ , namely,

$$(6.1) \quad \begin{aligned} \eta &= (\mathfrak{S}/24J)\xi^4 + \cdots, \\ \xi &= \frac{1}{6}\rho\xi^3 + (1/24)\rho(\mathfrak{M} + \mathfrak{S})\xi^4 \\ &\quad + (1/120)\rho[\mathfrak{M}_u - \mathfrak{S}_u + \mathfrak{M}^2 + \mathfrak{M}\mathfrak{S} - 2\mathfrak{S}^2 + 3\mathfrak{B} \\ &\quad + \frac{1}{2}(\mathfrak{M} - \mathfrak{S})(\mathfrak{M} - \phi_u)]\xi^5 + \cdots, \\ \tau &= \frac{1}{2}L\xi^2 + \frac{1}{6}L\mathfrak{S}\xi^3 + (1/24)L\mathfrak{B}\xi^4 \\ &\quad + (1/120)L\{\mathfrak{B}_u - (10/3)\mathfrak{S}(\mathfrak{S}_u + \mathfrak{S}^2) + (13/3)\mathfrak{B}\mathfrak{S} + 6\mathfrak{F} \\ &\quad + [\mathfrak{B} - (5/3)\mathfrak{S}^2](\mathfrak{M} - \phi_u)\}\xi^5 + \cdots. \end{aligned}$$

In like manner, the power series expansions of the v -curve in the neighborhood of the point x are

$$(6.2) \quad \begin{aligned} \xi &= (\mathfrak{T}/24I)\eta^4 + \cdots, \\ \xi &= \frac{1}{2}N\eta^2 + \frac{1}{6}N\mathfrak{R}\eta^3 + (1/24)N\mathfrak{C}\eta^4 \\ &\quad + (1/120)N\{\mathfrak{C}_v - (10/3)\mathfrak{R}(\mathfrak{R}_v + \mathfrak{R}^2) + (13/3)\mathfrak{C}\mathfrak{R} + 6\mathfrak{U} \\ &\quad + [\mathfrak{C} - (5/3)\mathfrak{R}^2](\mathfrak{R} - \psi_v)\}\eta^5 + \cdots, \\ \tau &= \frac{1}{6}\sigma\eta^3 + (1/24)\sigma(\mathfrak{D} + \mathfrak{R})\eta^4 \\ &\quad + (1/120)\sigma[\mathfrak{D}_v - \mathfrak{R}_v + \mathfrak{D}^2 + \mathfrak{D}\mathfrak{R} - 2\mathfrak{R}^2 + 3\mathfrak{C} \\ &\quad + \frac{1}{2}(\mathfrak{D} - \mathfrak{R})(\mathfrak{R} - \psi_v)]\eta^5 + \cdots. \end{aligned}$$

From the expansions (6.1), (6.2) it follows that the conditions that a general hyperquadric (5.1) has fourth order contact with both the u -, v -curves at the point x are

$$(6.3) \quad \begin{aligned} a_{33} &= (1/18N^2)[N(3\mathfrak{C} - 4\mathfrak{R}^2)a_{35} + \sigma(3\mathfrak{D} - \mathfrak{R})a_{45}], \\ a_{44} &= (1/18L^2)[\rho(3\mathfrak{M} - \mathfrak{S})a_{35} + L(3\mathfrak{B} - 4\mathfrak{S}^2)a_{45}]. \end{aligned}$$

This hyperquadric has fifth order contact with the u -curve at the point x in case

$$(6.4) \quad a_{34} = (1/10L)A_1a_{35} + (1/10\rho)A_2a_{45},$$

where we have placed

$$(6.5) \quad \begin{aligned} A_1 &= \mathfrak{X}_u - \mathfrak{S}_u + \mathfrak{X}^2 - (7/3)\mathfrak{X}\mathfrak{S} - (8/9)\mathfrak{S}^2 + (4/3)\mathfrak{B} \\ &\quad + \frac{1}{2}(\mathfrak{X} - \mathfrak{S})(\mathfrak{M} - \phi_u), \\ A_2 &= \mathfrak{B}_u - (10/3)\mathfrak{S}\mathfrak{S}_u + (10/9)\mathfrak{S}^3 - (2/3)\mathfrak{B}\mathfrak{S} + 6\mathfrak{F} \\ &\quad + [\mathfrak{B} - (5/3)\mathfrak{S}^2](\mathfrak{M} - \phi_u). \end{aligned}$$

Thus in the family of hyperquadrics having third order contact with the surface S at a point x , we obtain a pencil having fourth order contact with the v -curve and fifth order contact with the u -curve at the point x . In particular, a unique hyperquadric in this pencil may be determined to pass through the point y or z or to have fifth order contact with the v -curve at the point x .

7. Laplace transformed surfaces S_{-1} , S_1 . It is known that as u , v vary the Laplace transformed points x_{-1} , x_1 of an ordinary point x of the surface S with respect to the v -, u -curves of the conjugate net N_x generate two surfaces S_{-1} , S_1 , on which the parametric curves also form two conjugate nets N_{-1} , N_1 . As usual, we call the surfaces S_{-1} , S_1 and the nets N_{-1} , N_1 , respectively, the minus-first and first Laplace transformed surfaces and nets of N_x . In this section we shall first find the power series expansions of the surfaces S_{-1} , S_1 at the points x_{-1} , x_1 .

From system (1.2), equations (1.3), (2.8) and the substitution (1.5) by differentiation and substitution, any derivative of x_{-1} can be expressed as a linear combination of x , x_u , x_v , y , z . In particular, one obtains

$$(7.1) \quad \begin{aligned} x_{-1uu} &= (r - b_{uu} - bp)x + (m - 2b_u - b\alpha)x_u + \rho y + (\beta - bL)z, \\ x_{-1uv} &= (c_u - b_{uv} + ap)x + (a_u - b_v + c + a\alpha)x_u + aLz, \\ x_{-1vv} &= (c_v - b_{vv} + ac)x + (a_v + a^2)x_u + K\dot{x}_v, \\ x_{-1uuu} &= (*)x + (*)x_u + \rho Ex_v + \rho(\beta/L + \rho_u/\rho - N_u/N)y + (*)z, \\ x_{-1uuv} &= (*)x + (*)x_u + \rho y + (*)z, \\ x_{-1uvv} &= (*)x + (*)x_u + (K_u + bK)x_v + (*)z, \\ x_{-1vvv} &= (*)x + (*)x_u + [2K_v + K(a + \delta)]x_v + NKy, \end{aligned}$$

where $(*)$ denotes terms immaterial for our purpose.

The coordinates of any point X near the point x_{-1} on the surface S_{-1} can be represented by the Taylor's expansion as power series in the increments $\Delta u, \Delta v$ corresponding to displacement on S_{-1} from x_{-1} to the point X . From equations (7.1) we find that the local coordinates ξ_1, \dots, ξ_5 of the point X are given by the expansions

$$\begin{aligned}
 \xi_1 &= K\Delta v + \frac{1}{2}\mathfrak{F}\Delta u^2 + (K_u + bK)\Delta u\Delta v + \frac{1}{2}(K_v + 2aK)\Delta v^2 + \dots, \\
 \xi_2 &= 1 + (\alpha - b)\Delta u + a\Delta v + \frac{1}{2}(m - 2b_u - b\alpha)\Delta u^2 \\
 &\quad + (a_u - b_u + c + a\alpha)\Delta u\Delta v + \frac{1}{2}(a_v + a^2)\Delta v^2 + \dots, \\
 \xi_3 &= \frac{1}{2}K\Delta v^2 + \frac{1}{6}\rho E\Delta u^3 + \frac{1}{2}(K_u + bK)\Delta u\Delta v^2 \\
 &\quad + \frac{1}{6}[2K_v + K(a + \delta)]\Delta v^3 + \dots, \\
 \xi_4 &= \frac{1}{2}\rho\Delta u^2 + \frac{1}{6}\rho(\beta/L + \rho_u/\rho - N_u/N)\Delta u^3 + \frac{1}{2}a\rho\Delta u^2\Delta v \\
 &\quad + \frac{1}{6}NK\Delta v^3 + \dots, \\
 \xi_5 &= L\Delta u + \frac{1}{2}(\beta - bL)\Delta u^2 + aL\Delta u\Delta v + \dots.
 \end{aligned}
 \tag{7.2}$$

By means of the series (7.2) it follows that a hyperquadric with the general equation

$$\sum_{i,k=1}^5 a_{ik}\xi_i\xi_k = 0 \qquad (a_{ik} = a_{ki}),
 \tag{7.3}$$

has second order contact with the surface S_{-1} at the point x_{-1} in case

$$a_{22} = a_{25} = a_{12} = a_{15} = 0, \quad a_{55} = -(\rho/L^2)a_{24}, \quad a_{23} = -Ka_{11}.
 \tag{7.4}$$

Similarly, the hyperquadric (7.3) has second order contact with the surface S_1 at the point x_1 if, and only if,

$$a_{33} = a_{13} = a_{34} = a_{14} = 0, \quad a_{44} = -(\sigma/N^2)a_{35}, \quad a_{23} = -Ha_{11}.
 \tag{7.5}$$

Thus we reach the following theorem:⁴

Conjugate nets with equal and non-zero Laplace-Darboux invariants H, K in space S_4 are characterized by the property that there exists a proper hyperquadric (and therefore ∞^3 such hyperquadrics) having second order contact with both the Laplace transformed surfaces S_{-1}, S_1 at the Laplace transformed points x_{-1}, x_1 respectively.

Making use of the conditions (7.4) and the first three of the conditions (7.5) and imposing further on equation (7.3) the conditions that it be satisfied by the series (7.2) identically in terms of the third degree of $\Delta u, \Delta v$, it is easy to obtain the equations of a pencil of hyperquadrics having

⁴ This theorem is not true for conjugate nets in ordinary space. See the author's paper, "New geometrical characterizations of some special conjugate nets," *Duke Mathematical Journal*, vol. 12 (1945), p. 252.

third order contact with the surface S_{-1} at the point x_{-1} and first order contact with the surface S_1 at the point x_1 , namely,

$$(7.6) \quad \xi_1^2 - 2(\mathfrak{F}/\rho)\xi_1\xi_4 - 2K\xi_2\xi_3 - (1/N)[2K_v + K(\mathfrak{N} - \psi_v)]\xi_2\xi_4 \\ - (1/L)[2K_u + K(\mathfrak{M} - \phi_u)]\xi_3\xi_5 \\ - (1/3LN)\{(2\mathfrak{S} - \mathfrak{N})[2K_v + K(\mathfrak{N} - \psi_v)] - 2K\mathfrak{S}\}\xi_4\xi_5 \\ + (\rho/2L^2N)[2K_v + K(\mathfrak{N} - \psi_v)]\xi_5^2 + k_4\xi_4^2 = 0,$$

where k_4 is a parameter. There is also a pencil of hyperquadrics having third order contact with the surface S_1 at the point x_1 and first order contact with the surface S_{-1} at the point x_{-1} . The equation of this pencil can be obtained in a way similar to the foregoing, or else can be written immediately by making the substitutions (1.5), (2.7) and the necessary symmetrical interchanges of the subscripts; the result is

$$(7.7) \quad \xi_1^2 - 2(\mathfrak{G}/\sigma)\xi_1\xi_5 - 2H\xi_2\xi_3 - (1/N)[2H_v + H(\mathfrak{N} - \psi_v)]\xi_2\xi_4 \\ - (1/L)[2H_u + H(\mathfrak{M} - \phi_u)]\xi_3\xi_5 \\ - (1/3LN)\{(2\mathfrak{R} - \mathfrak{D})[2H_u + H(\mathfrak{M} - \phi_u)] - 2H\mathfrak{R}\}\xi_4\xi_5 \\ + (\sigma/2LN^2)[2H_u + H(\mathfrak{M} - \phi_u)]\xi_4^2 + k_5\xi_5^2 = 0,$$

where k_5 is a parameter. If a unique hyperquadric in the pencil (7.6) or (7.7) is desired, we may choose the one that passes through the point y or z . For these hyperquadrics we have $k_4 = 0$, $k_5 = 0$ respectively.

New geometric characterizations of the conjugate net N_x in some special cases may be deduced by studying the hyperquadrics (7.6), (7.7). In the first place, it is clear that the tangent plane at the point x of the surface S intersects the pencils (7.6), (7.7) in the conics of Koenigs, which coincide in case the conjugate net N_x has equal Laplace-Darboux invariants H , K . Two hyperquadrics (7.6), (7.7) with general k_4 , k_5 determine a pencil, in which there is one passing through the point x . The tangent hyperplanes of this hyperquadric at the points x , x_{-1} , x_1 are respectively

$$(7.8) \quad \sigma\mathfrak{F}\xi_4 - \rho\mathfrak{G}\xi_5 = 0,$$

$$(7.9) \quad (H - K)\xi_3 + (1/N)[H_v - K_v + \frac{1}{2}(H - K)(\mathfrak{N} - \psi_v)]\xi_4 = 0,$$

$$(7.10) \quad (H - K)\xi_2 + (1/L)[H_u - K_u + \frac{1}{2}(H - K)(\mathfrak{M} - \phi_u)]\xi_5 = 0.$$

The hyperplane (7.9) or (7.10) is indeterminate if, and only if, $H = K$. Thus we conclude that a necessary and sufficient condition that the hyperquadric passing through the point x in the pencil determined by any two hyperquadrics (7.6), (7.7) be a quadric hypercone with a vertex at the point x_{-1} or x_1 (and therefore with a line of vertices at $x_{-1}x_1$) is $H = K$.

The tangent hyperplanes of any hyperquadrics (7.6), (7.7) at the point x_{-1} are respectively

$$(7.11) \quad 2K\xi_3 + (1/N)[2K_v + K(\mathfrak{N} - \psi_v)]\xi_4 = 0,$$

$$(7.12) \quad 2H\xi_3 + (1/N)[2H_v + H(\mathfrak{N} - \psi_v)]\xi_4 = 0,$$

which coincide in case

$$(7.13) \quad H/K = U,$$

where U is a function of u alone. Moreover, the hyperplanes (7.11), (7.12) are separated harmonically by the hyperplanes $xx_{-1}yz$, $xx_{-1}x_1z$ if, and only if,

$$(7.14) \quad (\log HK)_v + \mathfrak{N} - \psi_v = 0.$$

Similarly, the tangent hyperplanes of any hyperquadrics (7.6), (7.7) at the point x_1 are respectively

$$(7.15) \quad 2K\xi_2 + (1/L)[2K_u + K(\mathfrak{M} - \phi_u)]\xi_5 = 0,$$

$$(7.16) \quad 2H\xi_2 + (1/L)[2H_u + H(\mathfrak{M} - \phi_u)]\xi_5 = 0,$$

which coincide in case

$$(7.17) \quad H/K = V,$$

where V is a function of v alone. Furthermore, the hyperplanes (7.15), (7.16) are separated harmonically by the hyperplanes xx_1yz , $xx_{-1}x_1y$ if, and only if,

$$(7.18) \quad (\log HK)_u + \mathfrak{M} - \phi_u = 0.$$

If the conditions (7.13), (7.17) hold simultaneously, then

$$(7.19) \quad H/K = \text{const.}$$

Moreover, if the conditions (7.14), (7.18) hold simultaneously, then the conjugate net N_x belongs to the class restricted by the condition

$$(7.20) \quad (\mathfrak{N} - \psi_v)_u = (\mathfrak{M} - \phi_u)_v.$$

Finally, we shall observe that in the pencil determined by any two hyperquadrics (7.6), (7.7) there is one whose tangent hyperplane at the point $x_{-1}(x_1)$ passes through the point $x_1(x_{-1})$. The polar hyperplane of the point x with respect to this hyperquadric is

$$(7.21) \quad (H - K)\xi_1 - (H\mathfrak{F}/\rho)\xi_4 + (K\mathfrak{G}/\sigma)\xi_5 = 0.$$

POLYNOMIAL MATRICES IN SEVERAL VARIABLES.*

By ERNST SNAPPER.

Introduction and Summary. Let $P[x_1, \dots, x_n]$ denote the ring of polynomials in n variables x_1, \dots, x_n whose coefficients belong to the commutative field P . The purpose of this paper is to develop the theory of an $s \times m$ matrix A whose elements belong to $P[x_1, \dots, x_n]$. The theory is a generalization of the case $n = 1$ which was treated in Part I of [1] (square brackets refer to the references) and is based on the module theory of Part II of [1]. In Part I of the present paper it is shown how a system S of partial, homogeneous, linear differential equations with constant coefficients in n independent variables and m dependent variables arises from A if the variables x_1, \dots, x_n are replaced by the differential operators $\partial/\partial t_1, \dots, \partial/\partial t_n$. The exponential solutions of S are then investigated by the use of the module theory of the row space of A . In Part II of this paper it is shown how the theory of the Hilbert characteristic function can be extended to a polynomial module M such as the row space or the column space of A . This extension gives rise to the notion of the *degrees* of M . The relationship between these degrees and the associated primes of M and the length of the primary components of M , as defined in [1], is derived. Furthermore, the norm and elementary divisor of M , as defined in [2], are studied. It is proved that the degrees of the primary factors of the norm of M are equal to the degrees of M and that, if the characteristic of P is zero, the multiplicities of the roots of the elementary divisor of M are equal to the p -exponents of a Noether decomposition of M . It is shown that the exponents which occur in the theory of the exponential solutions of S , as developed in Part I of this paper, are equal to the degrees and p -exponents of the row space of A and hence that, as in the case $n = 1$, the norm and the elementary divisor of the row space of A determine the algebraic properties of the exponential solutions of S . Finally, a system of algebraic, linear equations

$$(1) \quad \sum_{j=1}^m \alpha_{ij} Z_j = \gamma_i, \quad i = 1, \dots, s,$$

* Received October 4, 1945.

where α_{ij} and γ_i are given elements of $P[x_1, \dots, x_n]$, is studied. A criterion for the solvability of (1) by elements $Z_j \in P[x_1, \dots, x_n]$ is derived in terms of the degrees of the column space of A . The equivalence of this criterion with theorems 2.72 of [1] and 5.3 of [2] has been established.

I. Systems of Partial Differential Equations.

1.1. The vector space over the algebra of exponential and partial differentiation. This section is a generalization of Section 1 of [3] to systems of partial differential equations. Let $P[t_1, \dots, t_n]$ be a polynomial ring where P is a field of characteristic zero and where t_1, \dots, t_n are n indeterminates. Let L be the algebra of linear exponentials of $P[t_1, \dots, t_n]$ as defined in Section 1 of [3]. This means that L is the group ring with respect to $P[t_1, \dots, t_n]$ of the group H of linear exponentials. A linear exponential is a symbol $\exp(\alpha_1 t_1 + \dots + \alpha_n t_n)$ where $\alpha_i \in P$ and where the group H is defined by $\exp(\sum_{j=1}^n \alpha_j t_j) \exp(\sum_{j=1}^n \beta_j t_j) = \exp(\sum_{j=1}^n (\alpha_j + \beta_j) t_j)$. Let U be the m -dimensional column vector space over L as scalar domain. This means that U consists of the column vectors whose m components are elements of L . Since $\exp(\omega)$, where ω is the zero element of P , is the unit element of H , U contains as a submodule the m -dimensional column vector space W over $P[t_1, \dots, t_n]$ as scalar domain. The partial differential operators

$$\pi(D_1, \dots, D_n) = \sum_{j_1 \dots j_n} \gamma_{j_1 \dots j_n} D_1^{j_1} \dots D_n^{j_n}$$

where $\gamma_{j_1 \dots j_n} \in P$ are defined as in [3] for the elements $\xi \in L$. Hence, $\pi(D_1, \dots, D_n)(\xi)$ is computed by treating $D_1^{j_1} \dots D_n^{j_n}$ as the differential operator $\partial^{j_1+\dots+j_n}/\partial t_1^{j_1} \dots \partial t_n^{j_n}$ of analysis, the elements of P as constants

and the symbol $\exp(\sum_{j=1}^n \alpha_j t_j)$ as the exponential $e^{\alpha_1 t_1 + \dots + \alpha_n t_n}$ of analysis. Since the operators $\pi(D_1, \dots, D_n)$ can be added and multiplied as ordinary polynomials, these operators form a domain $P[D_1, \dots, D_n]$ which is isomorphic with the polynomial domain $P[x_1, \dots, x_n]$. Under this isomorphism, to every operator $\pi(D_1, \dots, D_n)$ there corresponds a polynomial $\pi(x_1, \dots, x_n)$ which is called the *auxiliary polynomial* of $\pi(D_1, \dots, D_n)$. We now consider the row vector operators $v(D_1, \dots, D_n)$ which are row vectors whose m components are elements of $P[D_1, \dots, D_n]$. If the components of $v(D_1, \dots, D_n)$ are π_1, \dots, π_m and if $u \in U$ has the components ξ_1, \dots, ξ_n , we define $v(D_1, \dots, D_n)(u) = \sum_{j=1}^m \pi_j(D_1, \dots, D_n)(\xi_j)$. Hence, the operators $v(D_1, \dots, D_n)$ transform vectors of U in elements of L , they can be added

and subtracted as ordinary row vectors with m components, and can be multiplied by operators of $P[D_1, \dots, D_n]$ in the usual way of scalar multiplication. Hence, the operators $v(D_1, \dots, D_n)$ form a vector space $V[D_1, \dots, D_n]$ over $P[D_1, \dots, D_n]$ as scalar domain which is isomorphic with the m -dimensional row vector space V over $P[x_1, \dots, x_n]$ as scalar domain. Under this isomorphism, to every operator vector $v(D_1, \dots, D_n) \in V[D_1, \dots, D_n]$ there corresponds a vector $v \in V$ which is called the *auxiliary vector* of $v(D_1, \dots, D_n)$. A system S of differential operators of $V[D_1, \dots, D_n]$ which is closed under vector subtraction and scalar multiplication is called an *operator module*. The auxiliary vectors of an operator module clearly form a module of V which is called the *auxiliary module* of S .

In the remainder of this section, S is a fixed operator module, M is its auxiliary module and \mathfrak{p} is the prime ideal $\mathfrak{p} = (x_1, \dots, x_n)$ of $P[x_1, \dots, x_n]$. As in [1], Sections 2.3 and 2.5, the module $M(\mathfrak{p})$ is defined as the isolated component of M all of whose associated primes are contained in \mathfrak{p} , and the module $M'(\mathfrak{p})$ as the isolated component of M all of whose associated primes are *properly* contained in \mathfrak{p} . A vector $w \in W$, i. e. a column vector whose m components are polynomials of $P[t_1, \dots, t_n]$, is called a *modal column* of S if $v(D_1, \dots, D_n)(w) = 0$ for all $v(D_1, \dots, D_n) \in S$. The term modal column is taken from the theory of systems of ordinary differential equations. (See [4], p. 179.) The theory of systems of partial differential equations of the next section is based on the following theory of modal columns.

Let S_k be the sub-operator module of S whose auxiliary module is $M \wedge \mathfrak{p}^k V$, i. e., S_k consists of those operators of S whose components have no derivatives of order less than k . (See [1], Section 2.1 for the definition of a product such as $\mathfrak{p}^k V$.) A column vector $w \in W$ is called *homogeneous of degree k* or a *vector form of degree k* if not all of its m components are zero and if every non-zero component of w is a homogeneous polynomial of degree k of $P[t_1, \dots, t_n]$. The degree of an arbitrary vector of W is of course the highest degree among its components. A vector w of degree k can be written uniquely as $w = \sum_{j=0}^k w_j$ where w_j is a vector form of degree j or is the zero vector and where $w_k \neq 0$.

LEMMA 1.11. Let $w = \sum_{j=0}^k w_j$ be the decomposition of the modal column w of degree k of S in homogeneous vector forms. Then, w_k is a modal column of S_k . Conversely, if w_k is a homogeneous modal column of degree k of S_k , we can find vector forms w_{k-1}, \dots, w_0 , where w_j has degree j or is the zero vector, such that $\sum_{j=0}^k w_j$ is a modal column of S .

Proof. Since the vector operators $v_a(D_1, \dots, D_n) \in S_a$ can be considered as linear functions on the vector space of vector forms w_a of degree α of W as $v_a(D_1, \dots, D_n)(w_a) \in P$, the above lemma is proved in exactly the same way as Lemma 1.1 of [3].

Let $S(\mathfrak{p})$ be the operator module of $V[D_1, \dots, D_n]$ whose auxiliary module is $M(\mathfrak{p})$.

LEMMA 1.12. *The vector w is a modal column of S if and only if w is a modal column of $S(\mathfrak{p})$.*

The proof of this lemma is the same as the proof of the more general Lemma 1.21 and hence is omitted here.

Now consider the sequence (1) $M(\mathfrak{p}) \subset M(\mathfrak{p}) : \mathfrak{p} \subset M(\mathfrak{p}) : \mathfrak{p}^2 \subset \dots$ where the symbol \subset , as everywhere else in this paper, denotes proper inclusion (see [1] Section 2.1 for the definition of quotients such as $M(\mathfrak{p}) : \mathfrak{p}^k$). Since the ascending chain condition holds in V (see [6], Section 99), the last term of this sequence is well defined and is the isolated component of $M(\mathfrak{p})$ whose associated primes do not contain \mathfrak{p} (see [5] Theorem 13). It follows that, since $M(\mathfrak{p}) = M'(\mathfrak{p}) \circ Q$ where Q is a primary component of M (see [1] Section 2.2) whose associated prime is \mathfrak{p} (or $M(\mathfrak{p}) = M'(\mathfrak{p})$ if \mathfrak{p} is not an associated prime of M) and since the associated primes of $M'(\mathfrak{p})$ are properly contained in \mathfrak{p} , the last term of sequence (1) is exactly $M'(\mathfrak{p})$. Furthermore, if ρ is the exponent of the fundamental ideal of Q (see [1] Section 2.1 for the definition of fundamental ideal), $M'(\mathfrak{p}) \circ \mathfrak{p}^\rho V \subseteq M(\mathfrak{p})$. The modal exponent δ of the operator module S is defined as the smallest integer such that $M'(\mathfrak{p}) \circ \mathfrak{p}^\delta V \subseteq M(\mathfrak{p})$.

We assert that, for any k , the factor module $M'(\mathfrak{p}) \circ \mathfrak{p}^k V / M'(\mathfrak{p}) \circ \mathfrak{p}^{k+1} V$ has finite P -rank, say σ_k . (P -rank means rank with respect to P . See [1] Section 1.1.) The P -rank of the factor module $\mathfrak{p}^k V / \mathfrak{p}^{k+1} V$ is clearly finite, namely precisely $m(n+k-1)!/k!(n-1)!$. Consequently, the maximal number of vectors of $M'(\mathfrak{p}) \circ \mathfrak{p}^k V$ which is P -linearly independent mod. $\mathfrak{p}^{k+1} V$ is finite and this number is the P -rank of $M'(\mathfrak{p}) \circ \mathfrak{p}^k V / M'(\mathfrak{p}) \circ \mathfrak{p}^{k+1} V$. In the same way, the P -rank of $M(\mathfrak{p}) \circ \mathfrak{p}^k V / M(\mathfrak{p}) \circ \mathfrak{p}^{k+1} V$ is finite, say τ_k .

LEMMA 1.13. *Always $\sigma_k \geq \tau_k$ and $\lambda_k = \sigma_k - \tau_k$ is always the P -rank of the factor module $(M(\mathfrak{p}), M'(\mathfrak{p}) \circ \mathfrak{p}^k V) / (M(\mathfrak{p}), M'(\mathfrak{p}) \circ \mathfrak{p}^{k+1} V)$. The factor module $M'(\mathfrak{p}) / M(\mathfrak{p})$ has finite P -rank $\lambda = \sum_{k=0}^{\delta-1} \lambda_k$, where δ is the modal exponent of S . If $k \geq \delta$, $\lambda_k = 0$ and hence $\sigma_k = \tau_k$.*

Proof. For the notion of sum of modules as $(M(\mathfrak{p}), M'(\mathfrak{p}) \circ \mathfrak{p}^k V)$, which

is the sum of the module $M(p)$ and the module $M'(p) \cap p^k V$, see [1] Section 2.1: From the sequence

$$(1) \quad M'(p) \cap p^\delta V \subset M'(p) \cap p^{\delta-1} V \subset \cdots \subset M'(p) \cap p V \subset M'(p)$$

and the fact that $M'(p) \cap p^\delta V \subseteq M(p)$, it follows that the maximal number of vectors of the module $M'(p) \cap p^k V$ which are P -linearly independent mod. $M(p)$ is finite, i. e., the P -rank of the module $M'(p) \cap p^k V / M(p) \cap p^k V$ is finite, say μ_k . Furthermore $(M(p), M'(p) \cap p^k V) / (M(p), M'(p) \cap p^{k+1} V)$ is isomorphic to M_1 / M_2 where $M_1 = (M(p), M'(p) \cap p^k V) / M(p)$ and $M_2 = (M(p), M'(p) \cap p^{k+1} V) / M(p)$. Since M_1 is isomorphic to $M'(p) \cap p^k V / M(p) \cap p^k V$ and M_2 is isomorphic to $M'(p) \cap p^{k+1} V / M(p) \cap p^{k+1} V$, the rank λ_k of M_1 / M_2 satisfies $\lambda_k = \mu_k - \mu_{k+1}$. It is seen immediately from the sequence (1) and the corresponding sequence for $M(p)$ that the rank of $M'(p) \cap p^k V / M(p) \cap p^{k+1} V$ is equal to both $\mu_k + \tau_k$ and $\mu_{k+1} + \sigma_k$ and hence that $\lambda_k = \sigma_k - \tau_k$. The sequence $M(p) = (M(p), M'(p) \cap p^\delta V) \subset (M(p), M'(p) \cap p^{\delta-1} V) \subset \cdots \subset (M(p), M'(p) \cap p V) \subset M'(p)$ then shows that the P -rank of $M'(p) / M(p)$ is finite and equal to $\lambda = \sum_{k=0}^{\delta-1} \lambda_k$. Finally, if $k \geq \delta$, $M'(p) \cap p^k V \subseteq M'(p) \cap p^\delta V \subseteq M(p)$ and hence $(M(p), M'(p) \cap p^k V) = M(p)$. Consequently, the P -rank of $(M(p), M'(p) \cap p^k V) / (M(p), M'(p) \cap p^{k+1} V)$ is then zero, i. e., $\lambda_k = 0$.

The modal rank λ of the operator module S is defined as the P -rank if $M'(p) / M(p)$. Let $S'(p)$ be the operator module of $V[D_1, \cdots, D_n]$ whose auxiliary module is $M'(p)$. Since $S(p) \subseteq S'(p)$, the modal columns of $S'(p)$ are also modal columns of $S(p)$, i. e., of S . A modal column of S which is at the same time a modal column of $S'(p)$ is called a *trivial modal column* of S ; otherwise the modal column is said to be *non-trivial*. These definitions of trivial and non-trivial will be justified by Lemma 1.22 and Remark 1.21.

For each k , let $v_{k1}, \cdots, v_{k\sigma_k}$ be σ_k vectors of $M'(p) \cap p^k V$ which are P -linearly independent mod. $M'(p) \cap p^{k+1} V$ and let $u_{k1}, \cdots, u_{k\tau_k}$ be τ_k vectors of $M(p) \cap p^k V$ which are P -linearly independent mod. $M(p) \cap p^{k+1} V$. Since $M(p) \cap p^k V \subseteq M'(p) \cap p^k V$ we can find, for each k , a $\tau_k \times \sigma_k$ matrix (α_{ij}) where $\alpha_{ij} \in P$, $i = 1, \cdots, \tau_k$, $j = 1, \cdots, \sigma_k$, such that

$$(1) \quad \begin{pmatrix} u_{k1} \\ \vdots \\ u_{k\tau_k} \end{pmatrix} = \begin{pmatrix} \alpha_{11} & \cdots & \alpha_{1\sigma_k} \\ \vdots & & \vdots \\ \alpha_{\tau_k 1} & \cdots & \alpha_{\tau_k \sigma_k} \end{pmatrix} \begin{pmatrix} v_{k1} \\ \vdots \\ v_{k\sigma_k} \end{pmatrix} \pmod{M'(p) \cap p^{k+1} V}.$$

The α_{ij} depend of course on k but the lower index k will be omitted for the α_{ij} 's. Each vector v_{kj} is the auxiliary vector of an operator $v_{kj}(D_1, \dots, D_n) \in S'(\mathfrak{p})$. Furthermore, the vectors v_{kj} ($j=1, \dots, \sigma_k$) are P -linearly independent mod. $\mathfrak{p}^{k+1}V$ since, otherwise, we could find constants $\beta_j \in P$ such that $\sum_{j=1}^{\sigma_k} \beta_j v_{kj} \in M'(\mathfrak{p}) \cap \mathfrak{p}^{k+1}V$. It follows easily from this and from the fact that $v_{kj}(D_1, \dots, D_n)$ has $m(n+k-1)!/k!(n-1)!$ derivatives of order k , that we can find $(m(n+k-1)!/k!(n-1)! - \sigma_k) = \mu_k$ P -linearly independent vector forms of degree k , $s_{k1}, \dots, s_{k\mu_k}$ where $s_{kj} \in W$, such that $v_{kj}(D_1, \dots, D_n)(s_{ki}) = 0$ for $j=1, \dots, \sigma_k$ and $i=1, \dots, \mu_k$. Since each vector of $M'(\mathfrak{p}) \cap \mathfrak{p}^k V$ is P -linearly dependent on $v_{k1}, \dots, v_{k\sigma_k}$ mod. $M'(\mathfrak{p}) \cap \mathfrak{p}^{k+1}V$, the vector forms $s_{k1}, \dots, s_{k\mu_k}$ are homogeneous modal columns of degree k of $S'_k(\mathfrak{p})$. According to Lemma 1.11 we can find μ_k modal columns $a_{k1}, \dots, a_{k\mu_k}$ of degree k of $S'(\mathfrak{p})$, i. e., trivial modal columns of S , where s_{kj} is the vector form of degree k in the decomposition of a_{kj} in vector forms. Each vector u_{kj} is the auxiliary vector of an operator $u_{kj}(D_1, \dots, D_n) \in S_k(\mathfrak{p})$ for $j=1, \dots, \tau_k$. Furthermore, the rows of the matrix (α_{ij}) are P -linearly independent since otherwise we could find constants $\beta_j \in P$ such that $\sum_{j=1}^{\tau_k} \beta_j u_{kj} \in M(\mathfrak{p}) \cap \mathfrak{p}^{k+1}V$; consequently, λ_k (see Lemma 1.12) is the rank of the null space of (α_{ij}) . It follows easily, from this and from the fact that $u_{k1}, \dots, u_{k\tau_k}$ are P -linearly independent mod. $\mathfrak{p}^{k+1}V$, that we can find λ_k P -linearly independent vector forms of degree k , $e_{k1}, \dots, e_{k\lambda_k}$, where $e_{kj} \in W$, such that $u_{kj}(D_1, \dots, D_n)(e_{ki}) = 0$ for $j=1, \dots, \tau_k$, $i=1, \dots, \lambda_k$ and where no e_{ki} satisfies all the σ_k operators $v_{kj}(D_1, \dots, D_n)$ ($j=1, \dots, \sigma_k$). Since every vector of $M(\mathfrak{p}) \cap \mathfrak{p}^k V$ is P -linearly dependent on $u_{k1}, \dots, u_{k\tau_k}$ mod. $M(\mathfrak{p}) \cap \mathfrak{p}^{k+1}V$ each e_{ki} is a homogeneous modal column of degree k of $S_k(\mathfrak{p})$. According to Lemma 1.11 we can find λ_k modal columns $f_{k1}, \dots, f_{k\lambda_k}$ of degree k of $S(\mathfrak{p})$, where e_{kj} is the vector form of degree k which occurs in the decomposition of f_{kj} in vector forms. Since e_{ij} is not a modal column of $S'_k(\mathfrak{p})$, it follows from Lemma 1.1 that f_{kj} is not a modal column of $S'(\mathfrak{p})$ and hence that $f_{k1}, \dots, f_{k\lambda_k}$ are λ_k non-trivial modal columns of $S(\mathfrak{p})$ and hence of S .

LEMMA 1.14. *The above vectors $f_{kj} \in W$ for $0 \leq k \leq \delta - 1$ and $1 \leq j \leq \lambda_k$ are λ non-trivial P -linearly independent modal columns of S , where δ is the modal exponent of S , λ the modal rank of S , and k is the degree of f_{kj} . An arbitrary modal column $w \in W$ of S is P -linearly dependent on the λ columns above and on a trivial modal column. If the degree of w*

is $k \leq \delta - 1$, w is P -linearly dependent on the $\sum_{q=0}^k \lambda_q$ modal columns f_{qj} for $0 \leq q \leq k$ and $1 \leq j \leq \lambda_q$ and on a trivial modal column.

Proof. We already know that the f_{kj} 's are non-trivial modal columns of S , that k is the degree of f_{kj} and that the number of f_{kj} 's, mentioned in Lemma 1.14, is $\sum_{k=0}^{\delta-1} \lambda_k$ which is the modal rank λ of S . Since e_{kj} is the vector form of degree k which occurs in the decomposition of f_{kj} in vector forms, the P -linear independence of the e_{kj} 's implies the P -linear independence of the λf_{kj} 's of Lemma 1.14. A vector form $w_k \in W$ of degree k is a modal column of $S_k(p)$ if and only if $u_{kj}(D_1, \dots, D_n)(w_k) = 0$ for $1 \leq j \leq \tau_k$. It follows from this fact and equation (1) that an arbitrary homogeneous modal column w_k of degree k of $S_k(p)$ is P -linearly dependent on the above $e_{k1}, \dots, e_{k\lambda_k}, s_{k1}, \dots, s_{k\mu_k}$. Let $w = \sum_{j=0}^k w_j$ be the decomposition of an arbitrary modal column w of degree k of S and hence of $S(p)$ in vector forms. According to Lemma 1.11, w_k is then a homogeneous modal column of degree k of $S_k(p)$ and hence

$$(2) \quad w_k = \sum_{j=1}^{\lambda_k} \alpha_j e_{kj} + \sum_{j=1}^{\mu_k} \beta_j s_{kj}.$$

If $k \geq \delta$, $\lambda_k = 0$ according to Lemma 1.13 and hence, in that case $w_k = \sum_{j=1}^{\mu_k} \beta_j s_{kj}$. Consider then the vector $g = f - \sum_{j=1}^{\mu_k} \beta_j a_{kj}$ where the a_{kj} 's have the same meaning as before and where hence $\sum_{j=1}^{\mu_k} \beta_j a_{kj}$ is a trivial modal column of S . Since g has degree at most $k-1$, it follows that f is the sum of a modal column of degree at most $k-1$ and a trivial modal column. Repetition of this process with g until the degree has become at most $\delta-1$ shows that every modal column of S is the sum of a modal column of degree at most $\delta-1$ and a trivial modal column. If $k \leq \delta-1$, equation (2) holds

and we consider $g = f - \sum_{j=1}^{\lambda_k} \alpha_j f_{kj} - \sum_{j=1}^{\mu_k} \beta_j a_{kj}$, where g is a modal column of S of degree at most $k-1$. Hence, f is then the sum of a modal column of degree at most $k-1$ of S , a modal column which is P -linearly dependent on $f_{k1}, \dots, f_{k\lambda_k}$, and a trivial modal column. Repetition of this process with g until the degree has become zero shows that f is the sum of a modal column h of degree zero, a modal column which is P -linearly dependent on f_{qj} for $1 \leq q \leq k$ and $1 \leq j \leq \lambda_k$, and a trivial modal column. Since the components of h are all elements of P , h either is the zero vector or is homogeneous and of degree zero and consequently P -linearly dependent on

$e_{01}, \dots, e_{0\lambda_0}, s_{01}, \dots, s_{0\lambda_0}$. Since $e_{0j} = f_{0j}$ for $1 \leq j \leq \lambda_0$, the lemma is proved.

1.2. Systems of partial differential equations. This section is an extension of Section 2 of [3] to m dependent variables and of Section 1.4 of [1] to n independent variables. A system of homogeneous, partial, linear differential equations with constant coefficients in m dependent variables and n independent variables is determined by a system S of operators of $V[D_1, \dots, D_n]$. (See 1.1.) A solution of S is defined as an element $z \in U$ such that $v(D_1, \dots, D_n)(z) = 0$ for all $v(D_1, \dots, D_n) \in S$. Here, U is the m -dimensional column vector space over M as scalar domain, where M is the algebra of linear exponentials of $K[t_1, \dots, t_n]$ and K is any extension field of P (see 1.1). From the isomorphism $V[D_1, \dots, D_n] \cong V$, where V is the m -dimensional row vector space over $P[x_1, \dots, x_n]$ as scalar domain (see 1.1), it follows that S generates an operator module. Since it is clear that this operator module and S have the same solutions, it is assumed hereafter that $S \subseteq V[D_1, \dots, D_n]$ is a fixed operator module and $M \subseteq V$ is its auxiliary module.

Let $w \exp(\xi t)$, be a vector of U , where (ξt) denotes the linear form $\sum_{j=1}^n \xi_j t_j$, $\xi_j \in K$, and where w is a vector of the m -dimensional column vector space W over $K[t_1, \dots, t_n]$ as scalar domain. If $v(D_1, \dots, D_n) \in V[D_1, \dots, D_n]$, then $v(D_1, \dots, D_n)(w \exp(\xi t))$ either is the zero element of M or is equal to $h \exp(\xi t)$ where $h \neq 0$ and $h \in P[t_1, \dots, t_n]$. Observe that, if $\pi(D_1, \dots, D_n) \in P[D_1, \dots, D_n]$ and h is a non-zero polynomial of $P[t_1, \dots, t_n]$, $\pi(D_1, \dots, D_n) h \exp(\xi t)$ cannot be zero unless $\pi(\xi_1, \dots, \xi_n) = 0$, where $\pi(\xi_1, \dots, \xi_n)$ is the value of the auxiliary polynomial $\pi \in P[x_1, \dots, x_n]$ of $\pi(D_1, \dots, D_n)$, for $x_j = \xi_j$ ($1 \leq j \leq n$). This follows from the fact that

$$\pi(D_1, \dots, D_n) h \exp(\xi t) = \pi(\xi_1, \dots, \xi_n) h \exp(\xi t) + g \exp(\xi t)$$

where $g \in P[t_1, \dots, t_n]$ is a polynomial whose degree is less than the degree of h .

In order to investigate the relationship of $w \exp(\xi t)$ to S , let $\mathfrak{p} \subseteq P[x_1, \dots, x_n]$ be the prime ideal which consists of the polynomials which vanish for $x_j = \xi_j$, $1 \leq j \leq n$. Again $M(\mathfrak{p})$ denotes the isolated component of M whose associated primes are contained in \mathfrak{p} and $M'(\mathfrak{p})$ the isolated component of M whose associated primes are properly contained in \mathfrak{p} . $S(\mathfrak{p})$ and $S'(\mathfrak{p})$ are the operator modules of $V[D_1, \dots, D_n]$ which have respectively $M(\mathfrak{p})$ and $M'(\mathfrak{p})$ as auxiliary modules.

LEMMA 1.21. *The vector $w \exp(\xi t)$, where $w \in W$, is a solution of S if and only if $w \exp(\xi t)$ is a solution of $S(\mathfrak{p})$.*

Proof. Since $S \subseteq S(\mathfrak{p})$, every solution of $S(\mathfrak{p})$ is a solution of S . Suppose that $w \exp(\xi t)$ is a solution of S . If $M = Q_1 \cap \cdots \cap Q_h$ is a Noether Decomposition of M and \mathfrak{p}_j is the radical of Q_j for $1 \leq j \leq h$, then $M(\mathfrak{p}) = Q_1 \cap \cdots \cap Q_m$ if $\mathfrak{p}_j \subseteq \mathfrak{p}$ for $1 \leq j \leq m$ and $\mathfrak{p}_j \not\subseteq \mathfrak{p}$ for $m < j \leq h$. Consequently, if ρ is the maximum of the exponents of the fundamental ideals of Q_{m+1}, \dots, Q_h , then $(\Pi_{j=m+1}^{\rho} \mathfrak{p}_j)^{\rho} M(\mathfrak{p}) \subseteq M$. Since $\mathfrak{p}_j \not\subseteq \mathfrak{p}$ for $m < j \leq h$, we can find $h - m$ polynomials $\pi_j \in P[x_1, \dots, x_n]$ such that $\pi_j \in \mathfrak{p}_j$ and $\pi_j(\xi_1, \dots, \xi_n) \neq 0$ for $m < j \leq h$. The polynomial $\pi = (\Pi_{j=m+1}^{\rho} \pi_j)^{\rho}$ then has the properties $\pi \in (\Pi_{j=m+1}^{\rho} \mathfrak{p}_j)^{\rho}$ and $\pi(\xi_1, \dots, \xi_n) \neq 0$. Let $\pi(D_1, \dots, D_n)$ be the operator of $V[D_1, \dots, D_n]$ whose auxiliary polynomial is π . If $w \exp(\xi t)$ is not a solution of $S(\mathfrak{p})$, there exists an operator $v(D_1, \dots, D_n) \in S(\mathfrak{p})$ such that $v(D_1, \dots, D_n) w \exp(\xi t) = g \exp(\xi t)$ where $g \in W$ and $g \neq 0$. The auxiliary operator v of $v(D_1, \dots, D_n)$ is a vector of $M(\mathfrak{p})$ and hence $\pi v \in M$ which implies $\pi(D_1, \dots, D_n) v(D_1, \dots, D_n) \in S$. However, $\pi(D_1, \dots, D_n) v(D_1, \dots, D_n) (w \exp(\xi t)) = \pi(D_1, \dots, D_n) (g \exp(\xi t))$ is not zero, since $g \neq 0$ and $\pi(\xi_1, \dots, \xi_n) \neq 0$ and hence $w \exp(\xi t)$ would not be a solution of S which is against the hypothesis. Hence, $w \exp(\xi t)$ is a solution of $S(\mathfrak{p})$ and the lemma is proved.

Since $S(\mathfrak{p}) \subseteq S'(\mathfrak{p})$, every solution of $S'(\mathfrak{p})$ is a solution of $S(\mathfrak{p})$, i. e. of S .

DEFINITION 1.21. *If a solution $w \exp(\xi t)$ of S , where $w \in W$, is also a solution of $S'(\mathfrak{p})$, the solution is said to be a trivial solution of S . Otherwise, $w \exp(\xi t)$ is called a non-trivial solution of S .*

Remark 1.21. This notion of triviality is an immediate extension of the notion of triviality used in [1], Section 1.41 and 1.61. If $n = 1$, $M'(\mathfrak{p})$ is always the closure of M , since then the non-zero prime ideals are always maximal.

We know from [1] that, if \mathfrak{p} is not an associated prime ideal of M , $M(\mathfrak{p}) = M'(\mathfrak{p})$. Hence then $S(\mathfrak{p}) = S'(\mathfrak{p})$ and all solutions $w \exp(\xi t)$ are trivial. In order to discuss the non-trivial solutions of S , the following assumptions are made for the remainder of this section: the ideal $\mathfrak{p} \subseteq P[x_1, \dots, x_n]$ is a fixed, $(n-i)$ -dimensional associated prime ideal of M ; ξ_1, \dots, ξ_n is a general point of \mathfrak{p} where ξ_j corresponds to the restclass of $x_j \bmod \mathfrak{p}$ for $1 \leq j \leq n$; the variables x_1, \dots, x_n are ordered in such a way that ξ_{i+1}, \dots, ξ_n are algebraically independent with respect to P and each ξ_j ,

for $1 \leq j \leq i$, is algebraic with respect to the field $\bar{P} = P(\xi_{i+1}, \dots, \xi_n)$; the vectors $\xi_1^{(j)}, \dots, \xi_i^{(j)}$ for $1 \leq j \leq c$, where $\xi_s^{(j)}$ is an element of the algebraic closure of \bar{P} , are the vectors which are conjugated to $\xi_1, \dots, \xi_i = \xi_1^{(1)}, \dots, \xi_i^{(1)}$ with respect to P ; P^{**} is the field which is obtained from P by adjunction to \bar{P} of all $\xi_s^{(j)}$ for $1 \leq j \leq c$, $1 \leq s \leq i$. We want to discuss those solutions $w^{**} \exp(\xi t)$ of S where (ξt) is $\xi_1 t_1 + \dots + \xi_n t_n$ and where the m components of the column vector w^{**} are polynomials of $P^{**}[t_1, \dots, t_i]$. The solution $w^{**} \exp(\xi t)$ is considered as a vector of the m -dimensional column vector space over the algebra of linear exponentials of $P^{**}[t_1, \dots, t_n]$ as scalar domain.

Remark 1.22. It may seem unnatural to allow only the variables t_1, \dots, t_i in the components of w^{**} , i.e., the variables which correspond to the algebraic coordinates of ξ_1, \dots, ξ_n . The justification is given by Section 3 of [3], where it is shown that, in the case of one dependent variable, this restriction of w^{**} arises in a natural way from the consideration of the initial conditions which may be imposed on the solutions of S .

If in each component of a vector $v \in M(p)$ we carry out the substitution $(\Xi)x_j = \xi_j$, $j = i+1, \dots, n$, this vector becomes a vector $\bar{v} \in \bar{V}$ where \bar{V} is the m -dimensional row vector space over $\bar{P}[x_1, \dots, x_i]$ as scalar domain. Let $\overline{M(p)}$ denote the module of \bar{V} which is generated by all vectors of $M(p)$ after the substitution (Ξ) has been carried out. Since $\bar{P} \subseteq P^{**}$, the vectors of $\overline{M(p)}$ can also be considered as vectors of the m -dimensional row vector space V^{**} over $P^{**}[x_1, \dots, x_i]$ as scalar domain. Let $M^{**}(p)$ denote the module of V^{**} which is generated by the vectors of $\overline{M(p)}$. Each component of each vector $v^{**} \in M^{**}(p)$ can be expanded with respect to the products $(x_1 - \xi_1)^{j_1} \dots (x_i - \xi_i)^{j_i}$, i.e., if v^{**} has the components $\pi^{**}_1, \dots, \pi^{**}_m$, where $\pi^{**}_k \in P^{**}[x_1, \dots, x_i]$ for $1 \leq k \leq m$, then

$$\pi^{**}_k = \sum_{j_1, \dots, j_i} a^{**}_{kj_1, \dots, j_i} (x_1 - \xi_1)^{j_1} \dots (x_i - \xi_i)^{j_i}$$

where $a^{**}_{kj_1, \dots, j_i} \in P^{**}$. If in each such expanded component we carry out the substitution $(D)(x_j - \xi_j) = D_j$, $j = 1, \dots, i$, we obtain an operator module $S^{**}_X(p)$ of $V^{**}[D_1, \dots, D_i]$ over $P^{**}[D_1, \dots, D_i]$ as scalar domain. Observe that the auxiliary module of $S^{**}_X(p)$ is not $M^{**}(p)$ but is the module $M^{**}_X(p)$ which arises from the expanded components of the vectors of $M^{**}(p)$ by the substitution $(X)x_j - \xi_j = X_j$, $j = 1, \dots, i$, where X_1, \dots, X_i are i determinates.

LEMMA 1.22. *The vector $w^{**} \exp(\xi t)$, where, as above, the components of w^{**} are polynomials of $P^{**}[t_1, \dots, t_i]$ and where ξ_1, \dots, ξ_n is a general*

point of the $(n-i)$ -dimensional associated prime \mathfrak{p} of M , is a solution of S if and only if w^{**} is a modal column of $S^{**}_X(\mathfrak{p})$. The solution $w^{**} \exp(\xi t)$ is a trivial solution of S if w^{**} is a trivial modal column of $S^{**}_X(\mathfrak{p})$ and is a non-trivial solution if w^{**} is a non-trivial modal column of $S^{**}_X(\mathfrak{p})$.

Proof. Since S and $S(\mathfrak{p})$ have the same solutions, let $v(D_1, \dots, D_n) \in S(\mathfrak{p})$ and let $v \in M(\mathfrak{p})$ be the auxiliary vector of $v(D_1, \dots, D_n)$. The vector v gives rise to an operator $v^{**}(D_1, \dots, D_i) \in S^{**}_X(\mathfrak{p})$, where $v^{**}(D_1, \dots, D_i)$ is obtained from v by the substitution $D \Xi$ (first Ξ and then D). These operators $v^{**}(D_1, \dots, D_i)$ generate $S^{**}_X(\mathfrak{p})$ and furthermore $v(D_1, \dots, D_n)(w^{**} \exp(\xi t)) = \exp(\xi t) v^{**}(D_1, \dots, D_i)(w^*)$ which is proved by direct computation in exactly the same way as Lemma 2.2 of [3]. Since $\exp(\xi t) v^{**}(D_1, \dots, D_i)(w^{**}) = 0$ if and only if $v^{**}(D_1, \dots, D_i)(w^*) = 0$, we have proved that $w^{**} \exp(\xi t)$ is a solution of $S(\mathfrak{p})$, i. e., of S , if and only if w^{**} is a modal column of $S^{**}_X(\mathfrak{p})$. In the same way we show that $w^{**} \exp(\xi t)$ is a solution of $S'(\mathfrak{p})$, i. e., a trivial solution of S , if and only if w^{**} is a modal column of $S'^{**}_X(\mathfrak{p})$, where $S'^{**}_X(\mathfrak{p})$ arises from $S'(\mathfrak{p})$ in the same way as $S^{**}_X(\mathfrak{p})$ from $S(\mathfrak{p})$. Since the auxiliary module of $S'^{**}_X(\mathfrak{p})$ is $M'^{**}_X(\mathfrak{p})$, where $M'^{**}_X(\mathfrak{p})$ arises from $M'(\mathfrak{p})$ as $M^{**}_X(\mathfrak{p})$ arises from $M(\mathfrak{p})$, all there remains to be shown is that $M'^{**}_X(\mathfrak{p})$ is the isolated component of $M^{**}_X(\mathfrak{p})$ whose associated primes are properly contained in the ideal $\mathfrak{X} = (X_1, \dots, X_i)$. For this purpose, we observe that $M(\mathfrak{p}) = \overline{M'}(\mathfrak{p}) \circ Q$ where Q is any primary component of M whose radical is \mathfrak{p} (see [1] Section 2.3). This implies that $\overline{M}(\mathfrak{p}) = \overline{M'}(\mathfrak{p}) \circ \overline{Q}$ where the bar again indicates the substitution Ξ (see [6], Section 97, where the same statement is proved for ideals). Since P^{**} is an algebraic extension of \bar{P} , we conclude that then $M^{**}(\mathfrak{p}) = \overline{M'}^{**}(\mathfrak{p}) \circ Q^{**}$. The module \overline{Q} is a primary module which has $\bar{\mathfrak{p}}$ as radical where $\bar{\mathfrak{p}}$ is the zero dimensional prime ideal of $\bar{P}[x_1, \dots, x_i]$ whose general point is ξ_1, \dots, ξ_i (see [6], Section 97, where the corresponding statement is proved for ideals; the proof can be immediately extended to modules). It follows that $Q^{**} = Q^{**}_1 \circ \dots \circ Q^{**}_c$ where Q^{**}_j is a primary module of V^{**} which has $\mathfrak{p}^{**}(\xi^{(j)}) = (x_1 - \xi_1^{(j)}, \dots, x_i - \xi_i^{(j)})$ as radical for $1 \leq j \leq c$ (see [7], Theorem 25; the proof of that theorem can immediately be extended to modules). Furthermore, from the fact that the associated primes of $M'(\mathfrak{p})$ are properly contained in \mathfrak{p} , it follows easily that the associated primes of $M'^{**}(\mathfrak{p})$ are properly contained in $\mathfrak{p}^{**}(\xi^{(j)})$, $1 \leq j \leq c$. Consequently, $M^{**}(\mathfrak{p}) = \overline{M'}^{**}(\mathfrak{p}) \circ Q^{**}_1 \circ \dots \circ Q^{**}_c$ and $M'^{**}(\mathfrak{p})$ is the isolated component of $M^{**}(\mathfrak{p})$ whose associated primes are properly contained in $\mathfrak{p}^{**}(\xi^{(1)})$. Since the substitution (X) establishes an isomorphism between the m -dimen-

sional row vector space V^{**} over $P^{**}[x_1, \dots, x_t]$ and the m -dimensional row vector space V^{**}_X over $P^{**}[X_1, \dots, X_t]$ which transforms $M^{**}(\mathfrak{p})$ in $M^{**}_X(\mathfrak{p})$, $M'^{**}(\mathfrak{p})$ in $M'^{**}_X(\mathfrak{p})$ and $\mathfrak{p}^{**}(\xi^{(1)})$ in \mathfrak{X} , $M'^{**}_X(\mathfrak{p})$ is the isolated component of $M^{**}_X(\mathfrak{p})$ whose associated primes are properly contained in \mathfrak{X} .

The modal rank $\lambda(\mathfrak{p})$ (see 1.1) of $S^{**}_X(\mathfrak{p})$ is the P^{**} -rank of the factor module $M'^{**}_X(\mathfrak{p})/M^{**}_X(\mathfrak{p})$. Because of the isomorphism (X) between V^{**}_X and V^{**} , $\lambda(\mathfrak{p})$ is also the P^{**} -rank of $M'^{**}(\mathfrak{p})/M^{**}(\mathfrak{p})$. The modal exponent $\delta(\mathfrak{p})$ (see 1.1) of $S^{**}_X(\mathfrak{p})$ is the smallest integer such that $M'^{**}_X(\mathfrak{p}) \cap \mathfrak{X}^{\delta(\mathfrak{p})} V^{**}_X \subseteq M^{**}_X(\mathfrak{p})$. If we call $\mathfrak{p}^{**}(\xi) = \mathfrak{p}^{**}(\xi^{(1)})$, the same isomorphism (X) shows that $\delta(\mathfrak{p})$ is the smallest integer such that $M'^{**}(\mathfrak{p}) \cap (\mathfrak{p}^{**}(\xi))^{\delta(\mathfrak{p})} V^{**} \subseteq M^{**}(\mathfrak{p})$.

DEFINITION 1.22. Let M be a module of the m -dimensional row vector space V over $P[x_1, \dots, x_n]$ as scalar domain. Let \mathfrak{p} be an associated prime of M with general point ξ_1, \dots, ξ_n where all notations and assumptions are as above. Then, the \mathfrak{p} -modal rank $\lambda(\mathfrak{p})$ of M is defined as the P^{**} -rank of $M'^{**}(\mathfrak{p})/M^{**}(\mathfrak{p})$. The \mathfrak{p} -modal exponent $\delta(\mathfrak{p})$ of M is defined as the smallest integer such that $M'^{**}(\mathfrak{p}) \cap (\mathfrak{p}^{**}(\xi))^{\delta(\mathfrak{p})} V^{**} \subseteq M^{**}(\mathfrak{p})$.

Let W^{**} denote the m -dimensional column vector space over $P^{**}[t_1, \dots, t_t]$ as scalar domain. The following theorem is a corollary of Lemmas 1.14 and 1.22.

THEOREM 1.21. There exist $\lambda(\mathfrak{p})$, P^{**} -linearly independent, non-trivial solutions $f^{**}_{kj} \exp(\xi t)$ of S . Here, $f^{**}_{kj} \in W^{**}$, the degree of f^{**}_{kj} is k where $0 \leq k \leq \delta(\mathfrak{p}) - 1$, $1 \leq j \leq \lambda_k(\mathfrak{p})$ and hence $\sum_{k=0}^{\delta(\mathfrak{p})-1} \lambda_k(\mathfrak{p}) = \lambda(\mathfrak{p})$. An arbitrary solution $w^{**} \exp(\xi t)$ of S , where $w^{**} \in W^{**}$, is P^{**} -linearly dependent on the $\lambda(\mathfrak{p})$ solutions above and on a trivial solution of S . If the degree of w^{**} is $k \leq \delta(\mathfrak{p}) - 1$, $w^{**} \exp(\xi t)$ is P^{**} -linearly dependent on the $\sum_{q=0}^k \lambda_q(\mathfrak{p})$ solutions $f^{**}_{qj} \exp(\xi t)$ for $0 \leq q \leq k$, $1 \leq j \leq \lambda_q(\mathfrak{p})$, and on a trivial solution of S .

Remark 1.23. If $m = 1$, Theorem 1.21 contains Theorem 2.1 of [3] as a corollary (in [3], $\delta(\mathfrak{p})$ is called the differential exponent instead of the \mathfrak{p} -modal exponent). If $n = 1$, Theorem 1.21 becomes Theorem 1.42 of [1]. Hence, the above theory of the exponential solutions of a system of partial differential equations and the theory of systems of linear algebraic equations of [1], Sections 2.6 and 2.7, tell us what exponents of an $s \times m$ polynomial

matrix $A = (\alpha_{ij})$, where $\alpha_{ij} \in P[x_1, \dots, x_n]$, we want to study. Namely, for each associated prime \mathfrak{p} of the row space of A (i. e., the module generated by the rows of A), we are interested in $\lambda(\mathfrak{p})$ and $\delta(\mathfrak{p})$; and, for each associated prime \mathfrak{p} of the column space of A , we are interested in the \mathfrak{p} -length $l(\mathfrak{p})$ (see [1] Section 2.6). The algebraic investigation of these exponents is based on the theory of Part II of [1] and on the theory of the Hilbert characteristic function of Part II of the present paper.

II. Polynomial Modules.

2.1. Homogeneous modules and their Noether decompositions. The theory of homogeneous modules is needed for the theory of the Hilbert characteristic functions of modules. Let P be an arbitrary field, not necessarily of characteristic zero, let V be the m -dimensional row vector space over $P[x_1, \dots, x_n]$ as scalar domain and let F be the m -dimensional row vector space over $P[x_0, x_1, \dots, x_n]$ as scalar domain. A vector $v \in F$ is called a *vector form* or *H-vector* of degree $\partial(v)$, if not all components of v are zero and if each of the non-zero components is a homogeneous polynomial, called a "form," of degree ∂ of $P[x_0, x_1, \dots, x_n]$. Elements of P are called constants and a constant vector is a vector whose components are elements of P . A non-zero constant vector is a vector form of degree zero and the zero vector is considered as a vector form of degree -1 . The degree $\partial(v)$ of an *arbitrary* vector v is the highest degree of its components and v can be written as a unique sum of vector forms $v = \sum_{j=0}^{\partial(v)} w_j$ where w_j is a vector form of degree j or the zero vector and where $w_{\partial(v)} \neq 0$. The vector forms w_j are called the *homogeneous components* of v .

DEFINITION 2.11. *The module $H \subseteq F$ is called a homogeneous module or H-module if the homogeneous components of any $v \in H$ are themselves vectors of H .*

Since every module of F has a finite number of generators, a module of F is homogeneous if and only if it can be generated by a finite number of vector forms.

We shall now consider the Noether decomposition of an H -module. Let L and M be two H -modules and \mathfrak{c} an H -ideal of $P[x_0, x_1, \dots, x_n]$ (see [7] Section 3). The sum (L, M) , the intersection $L \cap M$, the quotients $L : M$ and $M : \mathfrak{c}$, the product $\mathfrak{c}M$, the fundamental ideal $\mathfrak{f}(M)$ and the radical $\mathfrak{r}(M)$ are all defined in [1] as either modules of F or ideals of $P[x_0, \dots, x_n]$. It

can be shown easily that these modules and ideals are always H -modules and H -ideals. The *homogeneous kernel* M_H of an arbitrary module $M \subseteq F$ is defined as the module which is generated by the vector forms of M . Clearly M_H is the largest H -module which is contained in M and, if M is homogeneous, $M = M_H$. If L and M are two modules of F , it can be easily shown that $(L \circ M)_H = L_H \circ M_H$.

LEMMA 2.11. *Let M be a module of F . Then, $(f(M))_H = f(M_H)$ and $(r(M))_H = r(M)_H$. If Q is a primary module of F with \mathfrak{p} as radical, Q_H is primary and has \mathfrak{p}_H as radical.*

Proof. Since $M_H \subseteq M$, $f(M_H) \subseteq f(M)$ and, since we already know that $f(M_H)$ is homogeneous, $f(M_H) \subseteq (f(M))_H$. Now, let the form π of $P[x_0, \dots, x_n]$ be contained in $(f(M))_H$. Then $\pi \in f(M)$, which means that $\pi e_j \in M$ for $j = 1, \dots, m$ where e_j is the vector of F whose only non-zero component is the j -th one which is equal to unity. Since πe_j is a vector form we conclude that $\pi e_j \in M_H$ for $j = 1, \dots, m$, i. e. $\pi \in f(M_H)$. Consequently, all generators of $(f(M))_H$ are elements of $f(M_H)$ which proves that $(f(M))_H = f(M_H)$. Furthermore, $M_H \subseteq M$ implies, in the same way as for f , that $r(M_H) \subseteq (r(M))_H$. Let the form π of $P[x_0, \dots, x_n]$ be an element of $(r(M))_H$. Then, $\pi \in r(M)$ and hence, for some integer ρ , $\pi^\rho \in f(M)$. Since π^ρ is a form we conclude that $\pi^\rho \in (f(M))_H$ and hence that $\pi^\rho \in f(M_H)$ from which we conclude that $\pi \in r(M_H)$. Hence all generators of $(r(M))_H$ are contained in $r(M_H)$ which proves that $(r(M))_H = r(M_H)$. Let Q be a primary module of F with \mathfrak{p} as radical, then we know from the above that \mathfrak{p}_H is the radical of Q_H . It remains to be shown that, if $\pi \in P[x_0, \dots, x_n]$, $v \in F$ and $\pi v \in Q_H$, then either $\pi \in \mathfrak{p}_H$ or $v \in Q_H$. First suppose that π is a form and that $\sum_{j=0}^k w_j$ is the decomposition of v in vector forms. Then $\sum_{j=0}^k \pi w_j$ is the decomposition of πv and, since Q_H is homogeneous, we conclude that $\pi w_j \in Q_H$ for $0 \leq j \leq k$. Since $Q_H \subseteq Q$, either $\pi \in \mathfrak{p}$ and hence $\pi \in \mathfrak{p}_H$ or each $w_j \in Q$ and hence $w_j \in Q_H$ and the lemma is proved for the case that π is a form. If v is a vector form and π is an arbitrary polynomial, the lemma is proved in the same way. Consequently, we make the induction hypothesis that Lemma 2.11 is proved when either the number of homogeneous components of π is less than k or the number of homogeneous components of v is less than σ . Let $\pi = \sum_{j=0}^k \phi_j$ be the decomposition of π in forms and $v = \sum_{j=0}^\sigma w_j$ the decomposition of v in vector forms, where j is the degree of ϕ_j and of w_j and where $\phi_k \neq 0$, $w_\sigma \neq 0$. Then, $\phi_k w_\sigma$ is a homogeneous component

of πv , namely the one of highest degree, and hence $\phi_k w_\sigma \in Q_H$. If $\phi_k \notin \mathfrak{p}_H$, we conclude from the induction hypothesis that $w_\sigma \in Q_H$ and hence that $\pi \sum_{j=0}^{\sigma-1} w_j \in Q_H$. Since $\phi_k \notin \mathfrak{p}_H$ implies $\pi \notin \mathfrak{p}_H$, it follows from the induction hypothesis that $\sum_{j=0}^{\sigma-1} w_j \in Q_H$ and hence that $v = w_\sigma + \sum_{j=0}^{\sigma-1} w_j$ is contained in Q_H . If $\phi_k \in \mathfrak{p}_H$ and $v \notin Q_H$, we can find an integer ρ such that $\phi_k^\rho v \in Q_H$ and $\phi_k^{\rho-1} v \notin Q_H$ where ρ may be 1 ($\phi_k^0 v = v$). We then conclude that $(\sum_{j=0}^{k-1} \phi_j)(\phi_k^{\rho-1} v) \in Q_H$ and hence, from the induction hypothesis, that $\sum_{j=0}^{k-1} \phi_j \in \mathfrak{p}_H$. This makes $\pi = \phi_k + \sum_{j=0}^{k-1} \phi_j$ an element of \mathfrak{p}_H and the lemma is proved.

LEMMA 2.12. *Let M be an H -module of F . Then we can find a Noether decomposition $M = \bigcap_{j=1}^h Q_j$ of M where all Q_j for $1 \leq j \leq h$ are H -modules. Consequently, the associated primes of an H -module are homogeneous ideals.*

Proof. Let $M = \bigcap_{j=1}^h Q_j$ be any Noether decomposition of M . Since M is homogeneous, $M = M_H$, and, since the homogeneous kernel of an intersection is always the intersection of the kernels of the components, $M = \bigcap_{j=1}^h (Q_j)_H$. According to Lemma 2.11, $(Q_j)_H$ is primary and hence from $\bigcap_{j=1}^h (Q_j)_H$ we can obtain a Noether decomposition of M by grouping together those $(Q_j)_H$ which have the same radicals and by deleting the superfluous ones. Since every Noether decomposition of M must have exactly h components, it follows that $M = \bigcap_{j=1}^h (Q_j)_H$ itself is a Noether decomposition of M . Since $(Q_j)_H$ is a homogeneous module whose radical is hence homogeneous, Lemma 2.12 is proved.

Let \mathfrak{p} be a prime ideal of $P[x_0, \dots, x_n]$; let $M \subseteq F$; and let the \mathfrak{p} -closure $M(\mathfrak{p})$ be defined as in [1] Section 2.3. Then, it is in general not true that $(M(\mathfrak{p}))_H = M_H(\mathfrak{p}_H)$. For example, let $P[x_0, \dots, x_n] = P[x_0, x_1]$; let the dimension of F be one, i. e., $F = P[x_0, x_1]$; let M be the ideal $(x_0 - 1)$; and let \mathfrak{p} be the ideal (x_1) . Then, $M(\mathfrak{p}) = P[x_0, x_1]$ and hence $(M(\mathfrak{p}))_H = P[x_0, x_1]$. However, $M_H = (0)$; $\mathfrak{p}_H = (x_1)$; and hence $M_H(\mathfrak{p}_H) = (0)$. On the other hand, we assert that it is always true that, if M is \mathfrak{p} -closed (see [1] Section 2.3), M_H is \mathfrak{p}_H -closed. To say that M is \mathfrak{p} -closed is equivalent to saying that the associated primes of M , say $\mathfrak{p}_1, \dots, \mathfrak{p}_h$, are all contained in \mathfrak{p} . This implies that $(\mathfrak{p}_j)_H \subseteq \mathfrak{p}_H$ for $1 \leq j \leq h$. From the proof of Lemma 2.12, it follows that the associated primes of M_H are among the

$(\mathfrak{p}_j)_H$ for $1 \leq j \leq h$ and hence that they are all contained in \mathfrak{p}_H , which proves the assertion. If we take for \mathfrak{p} the zero ideal, \mathfrak{p}_H is also the zero ideal, and hence, if a module $M \subseteq F$ is closed (see [1] Section 2.7), then M_H is closed. Furthermore, if M is an H -module, Lemma 2.12 and the uniqueness of the isolated components of M (see [1] Section 2.2) imply that *isolated components of H -modules are always H -modules*. In particular, *all \mathfrak{C} -closures (see [1] Section 2.3) of a homogeneous module are always homogeneous modules*.

It is important to notice that in the present section we have considered the H -module M as imbedded in the complete vector space F . If $M \subseteq N$, where N is also an H -module of F , *this section undergoes no change whatsoever if M is considered as imbedded in N instead of in F* . In [1], the Noether decomposition of M as a submodule of any Noetherian vector space, not necessarily F , was considered. *The definitions, lemmas and proofs of this section remain valid when the Noetherian vector space N is considered as the imbedding space of M* .

2.2. H -modules of F and modules of V . The notation is the same as in the previous section. If $v \in F$, \bar{v} denotes the vector of V which arises from v by carrying out the substitution $(X_0)x_0 = 1$ in all the components of v . In the same way, the polynomial $\bar{\pi} \in P[x_1, \dots, x_n]$ arises from $\pi \in P[x_0, \dots, x_n]$ by carrying out (X_0) . Since $\overline{v_1 + v_2} = \bar{v}_1 + \bar{v}_2$ and $\bar{\pi v} = \pi \bar{v}$, (X_0) transforms a module $M \subseteq F$ in a module $\bar{M} \subseteq V$ and an ideal $\mathfrak{c} \subseteq P[x_0, \dots, x_n]$ in an ideal $\bar{\mathfrak{c}} \subseteq P[x_1, \dots, x_n]$ (see [7] Section 3 or [8] Section 4). As for ideals, it follows immediately that, if v_1, \dots, v_s are generators of M , $\bar{v}_1, \dots, \bar{v}_s$ are generators of \bar{M} and hence that $(\overline{M_1}, \overline{M_2}) = (\bar{M}_1, \bar{M}_2)$ and $\overline{\mathfrak{c}M} = \bar{\mathfrak{c}}\bar{M}$. If M is an H -module of F , the substitution (X_0) has to be carried out only in the vector forms of M in order to obtain \bar{M} (the proof is the same as for ideals, given in [7], p. 506). If $L \subseteq V$, L_0 always denotes the H -module of F which is generated by all the vector forms of F which are transformed by (X_0) in vectors of L . L_0 is called the *equivalent H -module* of L and is the largest H -module of F which is transformed in L by (X_0) .

LEMMA 2.21. *Let M be an H -module of F , \bar{M} the corresponding module of V , and M_0 the equivalent H -module of \bar{M} . Then, M_0 consists of all the vectors v of F for which there exists an integer $\rho \geq 0$ such that $x_0^\rho v \in M$. Hence, M_0 is the isolated component of M whose associated primes contain no power of x_0 .*

Proof. If v_1 and v_2 are vector forms of F where $\partial(v_1) \geq \partial(v_2)$, then $\bar{v}_1 = \bar{v}_2$ if and only if $v_1 = x_0^\rho v_2$ where $\rho = \partial(v_1) - \partial(v_2)$. It is clear that, if $v_1 = x_0^\rho v_2$, then $\bar{v}_1 = \bar{v}_2$; hence suppose that $\bar{v}_1 = \bar{v}_2$. Since the components of v_1 and v_2 are forms, it follows from [7], p. 507, that there exist non-negative integers ρ_1, \dots, ρ_m and $\sigma_1, \dots, \sigma_m$ such that

$$v_1 \begin{bmatrix} x_0^{\rho_1} \\ \vdots \\ x_0^{\rho_m} \end{bmatrix} = v_2 \begin{bmatrix} x_0^{\sigma_1} \\ \vdots \\ x_0^{\sigma_m} \end{bmatrix}.$$

This implies that $\sigma_i + \partial(v_2) = \rho_i + \partial(v_1)$ for each $1 \leq i \leq m$ and hence that $\partial(v_1) + \partial(v_2) = \sigma_i - \rho_i = \rho$. Consequently, $v_1 = x_0^\rho v_2$. The fact that M_0 consists of all vectors of F for which $x_0^\rho v \in M$ for some $\rho \geq 0$ can then be proved as for ideals (see [7] Theorem 8). In the terminology of [1], Section 2.3, this means that M_0 is the \mathbb{C} -closure of M where \mathbb{C} is the multiplicatively closed subset of $P[x_0, \dots, x_n]$ which consists of all powers of x_0 . The rest of Lemma 2.21 is a corollary of Lemma 2.31 of [1].

Let (T) denote a non-singular, linear transformation which transforms the variables x_0, \dots, x_n in the variables $y_0 = T(x_0), \dots, y_n = T(x_n)$. The transformation (T) establishes an isomorphism T between the m -dimensional row vector space F over $P[x_0, \dots, x_n]$ and the m -dimensional row vector space $T(F)$ over $P[y_0, \dots, y_n]$. If $v \in F$ then $T(v)$ is the vector of $T(F)$ which arises from v by carrying out the substitution (T) in the components of v . Since T is an isomorphism, a module $M \subseteq F$ is transformed in a module $T(M) \subseteq T(F)$ and, if $\mathfrak{p}_1, \dots, \mathfrak{p}_h$ are the associated primes of M , $T(\mathfrak{p}_1), \dots, T(\mathfrak{p}_h)$ are the associated primes of $T(M)$. If $\mathfrak{p}_1, \dots, \mathfrak{p}_{h-1}$ are all different from the ideal (x_0, \dots, x_n) and P is infinite, (T) can be chosen in such a way that no power x_0^ρ is contained in any $T(\mathfrak{p}_j)$ for $1 \leq j \leq h-1$ (see [7] p. 507). This proves the following lemma.

LEMMA 2.22. *If P is infinite, after a suitable non-singular linear transformation the M_0 of Lemma 2.21 is the isolated component of the M of Lemma 2.21 whose associated primes are different from (x_0, \dots, x_n) .*

In other words, after a suitable linear transformation, either $M_0 = M$, or M has a primary component whose radical is (x_0, \dots, x_n) , in which case M_0 is obtained from a Noether decomposition of M by omitting that primary component.

In the same way, the remainder of [7], Section 3, and [8], part I, can be extended to polynomial modules. We shall use the following lemma whose proof, being a direct extension of the proof for the corresponding lemma for ideals, is omitted. For any module $M \subseteq V$ and prime ideal $\mathfrak{p} \subseteq P[x_1, \dots, x_n]$, the modules $M'(\mathfrak{p})$ and $M(\mathfrak{p})$ are defined as in [1], Sections 2.3 and 2.5, and the lower index 0 always denotes the equivalent H -module. In the same way, if \mathfrak{C} is any multiplicatively closed subset of $P[x_1, \dots, x_n]$ which does not contain the zero element, $M(\mathfrak{C})$ is defined in [1] Section 2.3 and \mathfrak{C}_0 denotes the multiplicatively closed subset of all forms of $P[x_0, \dots, x_n]$ which are transformed in polynomials of \mathfrak{C} by the substitution (X_0) .

LEMMA 2.23. Let $M = \bigcap_{j=1}^h Q_j$ be a Noether decomposition of the module $M \subseteq V$ where \mathfrak{p}_j is the radical of Q_j . Then, $M_0 = \bigcap_{j=1}^h (Q_{j0})$ is a Noether decomposition of M_0 and \mathfrak{p}_{j0} is the radical of Q_{j0} . It follows that, if Q is a primary (or prime) module of V with \mathfrak{p} as radical, Q_0 is a primary (or prime) module of F with \mathfrak{p}_0 as radical. It furthermore follows that always $(M'(\mathfrak{p}))_0 = M'_0(\mathfrak{p}_0)$, $(M(\mathfrak{p}))_0 = M_0(\mathfrak{p}_0)$ and $(M(\mathfrak{C}))_0 = M_0(\mathfrak{C}_0)$.

For any H -module M of F , the reduced dimension $d(M)$ is the largest reduced dimension of its non-zero associated primes. If M has only the zero ideal of $P[x_0, \dots, x_n]$ as associated prime, i. e. if M is closed (see [1] Section 2.7), $d(M)$ is considered as -1 . Since, according to Lemma 2.12, the associated primes of M are H -ideals of $P[x_0, \dots, x_n]$, the above definition of $d(M)$ is complete (see [7] p. 510 for the definition of reduced dimension of an H -ideal).

Again, it is important to observe that in the present section we have used as the imbedding Noetherian vector space of an H -module M , the vector space F , and of a non-homogeneous module \bar{M} the vector space V . If N is any homogeneous module of F where $N = N_0$ and $M \subseteq N$, all definitions, lemmas and proofs of this section remain valid when N is considered as the imbedding Noetherian factor space of M and \bar{N} as the imbedding space of \bar{M} .

2.3. The Hilbert characteristic function of a polynomial module.

In this section it is shown how the work of B. L. v. d. Waerden on the Hilbert characteristic function (see [7], Section 4 and [8], part II) can be extended to polynomial modules. The motivation for this study is the fact that it enables us to investigate the exponents $\lambda(\mathfrak{p})$, $\delta(\mathfrak{p})$ and $l(\mathfrak{p})$ (see Remark 1.23). The notation is the same as in the previous sections. P -linear

dependence and P -linear independence means linear dependence and independence with respect to P . There exists a one to one correspondence between the vector forms of degree ρ of F and the vectors of the $m(n + \rho - 1)!/n!(\rho - 1)!$ -dimensional row vector space Φ over P as scalar domain. Since this correspondence is an operator isomorphism with respect to P as operator domain, the notion of P -linear dependence for vector forms of degree ρ of F follows the rules of P -linear dependence for vectors of Φ . If M is an H -module of F , $\phi(\rho; M)$ denotes the maximum number of vector forms of degree ρ of M which are P -linearly independent. It follows from the above remark that, for two H -modules L and M ,

$$\phi(\rho; (L, M)) = \phi(\rho; L) + \phi(\rho; M) - \phi(\rho; L \circ M)$$

(designated equation I). If $M \subseteq N$ are two H -modules of F , $\chi(\rho; N/M)$ denotes the maximum number of vector forms of degree ρ of N which are P -linearly independent mod. M . As above, we conclude that

$$\chi(\rho; N/M) = \phi(\rho; N) - \phi(\rho; M)$$

(designated equation II). From equations I and II, it follows that, if $(L, M) \subseteq N$ where L is also an H -module of F , then

$$\chi(\rho; N/(L, M)) = \chi(\rho; N/L) + \chi(\rho; N/M) - \chi(\rho; N/L \circ M)$$

(designated equation III see [7] p. 511 for the corresponding proof for ideals).

LEMMA 2.31. Let $\psi \in P[x_0, \dots, x_n]$ be a form of degree γ and $L \subseteq M$ be two H -modules of F . Let ψ be M -relatively prime to L , i. e., if $\psi v \in L$ where v is a vector form of M then $v \in L$. Then,

$$\chi(\rho; M/(L, \psi M)) = \chi(\rho; M/L) - \chi(\rho - \gamma; M/L).$$

Proof. The product ψM of a form and a module is defined in [1] Section 2.1. If v is a vector form of degree ρ of ψM , then $v = \psi m$ where m is a vector form of degree $\rho - \gamma$ of M and hence $\phi(\rho; \psi M) = \phi(\rho - \gamma; M)$. If v is a vector form of degree ρ of $L \circ \psi M$, then $v = \psi m$ where m is a vector form of degree $\rho - \gamma$ of M and where, furthermore, $\psi m \in L$ and hence $m \in L$. It follows that $\phi(\rho; L \circ \psi M) = \phi(\rho - \gamma; L)$. Then, from equation III, it follows that

$$\chi(\rho; M/(L, \psi M)) = \chi(\rho; M/L) + \chi(\rho; M/\psi M) - \chi(\rho; M/L \circ \psi M)$$

and, from equation II, it follows that $\chi(\rho; M/\psi M) = \phi(\rho; M) - \phi(\rho; \psi M)$ and $\chi(\rho; M/L \circ \psi M) = \phi(\rho; M) - \phi(\rho; L \circ \psi M)$. Consequently,

$$\chi(\rho; M/(L, \psi M)) = \chi(\rho; M/L) + \phi(\rho - \gamma; L) - \phi(\rho - \gamma; M)$$

and hence according to equation II,

$$\chi(\rho; M/(L, \psi M)) = \chi(\rho; M/L) - \chi(\rho - \gamma; M/L).$$

As in [1], Section 2.3, let \mathfrak{C} denote a multiplicatively closed subset of $P[x_0, \dots, x_n]$ which does not contain the zero element ω of $P[x_0, \dots, x_n]$ and let $M(\mathfrak{C})$ denote the \mathfrak{C} -closure of the H -module M of F . Since $M(\mathfrak{C})$ is an isolated component of M or is equal to F , a Noether decomposition of M gives rise to a decomposition $M = M(\mathfrak{C}) \cap Q_1 \cap \dots \cap Q_h$ where the radical \mathfrak{p}_j of Q_j , $1 \leq j \leq h$, is an H -ideal of $P[x_0, \dots, x_n]$ which contains an element of \mathfrak{C} . According to [1], Lemma 2.21, if $Q_j^* = M(\mathfrak{C}) \cap Q_j$ for $1 \leq j \leq h$, $M = Q_1^* \cap \dots \cap Q_h^*$ is a Noether decomposition of M as a submodule of the Noetherian vector space $M(\mathfrak{C})$ and the radical of Q_j^* as a primary submodule of $M(\mathfrak{C})$ is \mathfrak{p}_j .

LEMMA 2.32. *Let M be an H -module of F and let the reduced dimension d of M as a submodule of the Noetherian vector space $M(\mathfrak{C})$ be -1 . Then, for large enough ρ , $\chi(\rho; M(\mathfrak{C})/M) = 0$.*

Proof. Let, as above, $M = M(\mathfrak{C}) \cap Q_1 \cap \dots \cap Q_h$. Since the associated primes of M , considered as a submodule of $M(\mathfrak{C})$, are $\mathfrak{p}_1, \dots, \mathfrak{p}_h$, h is at most 2 and $\mathfrak{p}_1 = (x_0, \dots, x_n)$ and $\mathfrak{p}_2 = (\omega)$. However, if $\mathfrak{p}_2 = (\omega)$, $M(\mathfrak{C})$ could not be isolated or equal to F (see [1] Sections 2.3 and 2.7) and hence $h = 1$ and $\mathfrak{p}_1 = (x_0, \dots, x_n)$. It follows that, for large enough ρ ,

$$\mathfrak{p}_1^\rho F \cap M(\mathfrak{C}) \subseteq Q_1 \cap M(\mathfrak{C}) = M$$

which means that for large enough ρ the vector forms of degree ρ of $M(\mathfrak{C})$ are contained in M and hence that $\chi(\rho; M(\mathfrak{C})/M) = 0$.

THEOREM 2.31. *Let M be an H -module of F whose reduced dimension as a submodule of the Noetherian vector space $M(\mathfrak{C})$ is d . Then, for large enough ρ , $\chi(\rho; M(\mathfrak{C})/M) = a_0 \binom{\rho}{d} + a_1 \binom{\rho}{d-1} + \dots + a_d$, where $\binom{\rho}{d-j} = \rho!/(d-j)!(\rho-d+j)!$ and where a_0, \dots, a_d are whole numbers independent of ρ .*

Proof. As in the proof of [7], Theorem 17, we observe that, when $d = -1$, Theorem 2.31 follows from Lemma 2.32. Hence, we make the induction hypothesis that Theorem 2.31 has been proved for $d = -1, 0, \dots, d-1$ and we assume that the H -module M has reduced dimension d as a submodule of $M(\mathfrak{C})$. We first consider the case when M is primary as a

submodule of $M(\mathbb{C})$, i. e., when $M = M(\mathbb{C}) \cap Q$ where \mathfrak{p} is the associated prime of Q . Then \mathfrak{p} cannot be (ω) since $M(\mathbb{C})$ could then be neither isolated nor equal to F . Furthermore, $(x_0, \dots, x_n) \not\subseteq \mathfrak{p}$ for, since (x_0, \dots, x_n) is a maximal prime ideal, this would imply that $\mathfrak{p} = (x_0, \dots, x_n)$ and hence that $d = -1$. Consequently, there exists a form $\psi \in P[x_0, \dots, x_n]$ of degree one where $\psi \in \mathfrak{p}$. Since $\psi v \in Q$ implies $v \in Q$ we conclude, from Lemma 2.31, that

$$\chi(\rho; M(\mathbb{C})/(M, \psi M(\mathbb{C}))) = \chi(\rho; M(\mathbb{C})/M) - \chi(\rho - 1; M(\mathbb{C})/M)$$

(designated equation IV). We assert that, if $L = (M, \psi M(\mathbb{C}))$, then $L(\mathbb{C}) = M(\mathbb{C})$ and the reduced dimension of L as a submodule of the Noetherian vector space $M(\mathbb{C})$ is less than d . Since $M \subseteq L$, $M(\mathbb{C}) \subseteq L(\mathbb{C})$; and, since $L \subseteq M(\mathbb{C})$ and $M(\mathbb{C})$ is \mathbb{C} -closed (see [1] Lemma 2.31), $L(\mathbb{C}) \subseteq M(\mathbb{C})$ which proves the first part of the assertion. Consequently we can again write $L = M(\mathbb{C}) \cap Q'_1 \cap \dots \cap Q'_k$ where Q'_j is a primary module of F which has \mathfrak{p}'_j as radical, and where, as before, $\mathfrak{p}'_1, \dots, \mathfrak{p}'_k$ are the associated primes of L as a submodule of $M(\mathbb{C})$. Since $\psi M(\mathbb{C}) \subseteq L \subseteq Q'_j$ and $M(\mathbb{C}) \not\subseteq Q'_j$, $\psi \in \mathfrak{p}'_j$ for $1 \leq j \leq k$. Since $M = M(\mathbb{C}) \cap Q$ and since, for large enough ρ , $\mathfrak{p}^\rho M(\mathbb{C}) \subseteq M(\mathbb{C}) \cap Q$, we conclude that, for large enough ρ , $\mathfrak{p}^\rho M(\mathbb{C}) \subseteq L \subseteq Q'_j$ and hence $\mathfrak{p} \subseteq \mathfrak{p}'_j$ for $1 \leq j \leq k$. Hence, $\mathfrak{p} \subseteq (\mathfrak{p}, \psi) \subseteq \mathfrak{p}'_j$ for $1 \leq j \leq k$ which shows that the reduced dimension of each \mathfrak{p}'_j is less than d and the assertion is proved. We then conclude from the induction hypothesis that for large enough ρ

$$\chi(\rho; M(\mathbb{C})/(M, \psi M(\mathbb{C}))) = a_0 \binom{\rho}{d-1} + a_1 \binom{\rho}{d-2} + \dots + a_{d-2} \binom{\rho}{1} + a_{d-1}$$

where a_0, \dots, a_{d-1} are whole numbers. If this expression is substituted in equation IV, we obtain a recursion formula for $\chi(\rho; M(\mathbb{C})/M)$ from which, exactly as in [7], p. 513. it follows that

$$\chi(\rho; M(\mathbb{C})/M) = a_0 \binom{\rho}{d} + A_1 \binom{\rho}{d-1} + \dots + A_{d-1} \binom{\rho}{1} + A_d$$

where A_1, \dots, A_d are whole numbers and where a_0 has the same meaning as above and hence is non-negative. Hence, Theorem 2.31 is proved for the special case when $h = 1$ in the decomposition $M = M(\mathbb{C}) \cap Q_1 \cap \dots \cap Q_h$. We now make the induction hypothesis that Theorem 2.31 has been proved, not only for $d = -1, 0, \dots, d-1$, but also for $h = 1, 2, \dots, h-1$. We then observe that $M = L \cap N$ where

$$L = M(\mathbb{C}) \cap Q_1 \cap \dots \cap Q_{k-1} \cap Q_{k+1} \cap \dots \cap Q_h \text{ and } N = M(\mathbb{C}) \cap Q_k$$

and where Q_k is any one of the primary components Q_1, \dots, Q_h whose

reduced dimension is exactly d . From equation III we conclude that (1) $\chi(\rho; M(\mathfrak{C})/M) = \chi(\rho; M(\mathfrak{C})/L) + \chi(\rho; M(\mathfrak{C})/N) - \chi(\rho; M(\mathfrak{C})/(L, N))$. We then prove as before that $L(\mathfrak{C}) = N(\mathfrak{C}) = (L, N)(\mathfrak{C}) = M(\mathfrak{C})$ and that the reduced dimension of (L, N) as a submodule of the Noetherian vector space $M(\mathfrak{C})$ is less than d . Consequently, according to the induction hypothesis, Theorem 2.31 is proved for each term of the right hand side of (1) and hence the theorem is proved for $\chi(\rho; M(\mathfrak{C})/M)$.

The polynomial $\chi(\rho; M(\mathfrak{C})/M) = \sum_{j=0}^d a_j \binom{\rho}{d-j}$ will be called the *characteristic function of M with respect to $M(\mathfrak{C})$* . The coefficient a_0 will be called the *degree of M with respect to $M(\mathfrak{C})$* . We shall talk about the $M(\mathfrak{C})$ -characteristic function, the $M(\mathfrak{C})$ -degree and the $M(\mathfrak{C})$ -reduced dimension of M .

THEOREM 2.32. *If, as before, $M = M(\mathfrak{C}) \cap Q_1 \cap \cdots \cap Q_h$ and d is the $M(\mathfrak{C})$ -reduced dimension of the H -module M , then the $M(\mathfrak{C})$ -degree of M is the sum of the $M(\mathfrak{C})$ -degrees of those components $M(\mathfrak{C}) \cap Q_k$ where the radical \mathfrak{p}_k of Q_k has reduced dimension d .*

Proof. We use the notation of the proof of Theorem 2.31. Since the $M(\mathfrak{C})$ -reduced dimension of (L, N) is less than d , the $M(\mathfrak{C})$ -degree of M is equal to the sum of the $M(\mathfrak{C})$ -degrees of L and N , according to equation (1). Since N is already of the form $M(\mathfrak{C}) \cap Q_k$, where Q_k has reduced dimension d , we have to prove Theorem 2.32 only for L . Since L , considered as a submodule of $M(\mathfrak{C})$, has only $h-1$ components, Theorem 2.32 is proved by induction as soon as the theorem is proved for $h=1$. However, for $h=1$ Theorem 2.32 is trivial.

THEOREM 2.33. *If the $M(\mathfrak{C})$ -reduced dimension d of the H -module M satisfies $d > -1$, the $M(\mathfrak{C})$ -degree a_0 of M satisfies $a_0 > 0$.*

Proof. If $d=0$, according to Theorem 2.31 $\chi(\rho; M(\mathfrak{C})/M) = a_0$. Then, $a_0=0$ would imply that for large enough ρ all the vector forms of degree ρ of $M(\mathfrak{C})$ are contained in M , i.e., that $\mathfrak{p}^\rho F \cap M(\mathfrak{C}) \subseteq M$ where $\mathfrak{p} = (x_0, \dots, x_n)$. We would conclude that $\mathfrak{p}^\rho M(\mathfrak{C}) \subseteq M$ and hence, as in the proof of Theorem 2.31, that \mathfrak{p} is contained in every associated prime of M as a submodule of $M(\mathfrak{C})$. It would follow that $M = M(\mathfrak{C}) \cap Q_1$ where \mathfrak{p} is the radical of the primary module Q_1 and hence that $d=-1$. Hence, when $d=0$, $a_0 \neq 0$ and, since a_0 represents a number of vectors, $a_0 > 0$. Hence, we assume that Theorem 2.33 has been proved for $M(\mathfrak{C})$ -reduced dimensions $0, 1, \dots, d-1$ and that the $M(\mathfrak{C})$ -reduced dimension of M is d .

It follows from Theorem 2.32 that Theorem 2.33 has to be proved only for the case $h = 1$. According to the proof of Theorem 2.31, if $h = 1$, the $M(\mathfrak{C})$ -degree of M is equal to the coefficient of $\binom{\rho}{d-1}$ of the characteristic function $\chi(\rho; M(\mathfrak{C})/(M, \psi M(\mathfrak{C})))$. Hence, according to the induction hypothesis, Theorem 2.33 is proved as soon as we have proved that the $M(\mathfrak{C})$ -reduced dimension of $(M, \psi M(\mathfrak{C}))$ is exactly $d - 1$. The proof of this fact is an immediate extension of [7], p. 514, to modules.

Let M be any module of the m -dimensional row vector space V over $P[x_1, \dots, x_n]$ as scalar domain. Let \mathfrak{C} be a multiplicatively closed subset of $P[x_1, \dots, x_n]$ which does not contain the zero element and let M have dimension d as a sub-module of the Noetherian vector space $M(\mathfrak{C})$. This means that, in the representation $M = M(\mathfrak{C}) \cap Q_1 \cap \dots \cap Q_h$ where, as before, Q_j is a primary module of V with \mathfrak{p}_j as radical, the highest dimension of the prime ideals $\mathfrak{p}_1, \dots, \mathfrak{p}_h$ is d . We know that then $M_0 = M_0(\mathfrak{C}_0) \cap Q_{10} \cap \dots \cap Q_{h_0}$ where the lower index zero denotes the equivalent H -modules. Since the substitution (X_0) establishes a one to one correspondence between the vector forms of degree ρ of M_0 and the vectors of degree $\leq \rho$ of M , the following theorem is a corollary of Theorem 2.31, 2.32 and 2.33.

THEOREM 2.34. *Let the arbitrary module M of V have the dimension d as a proper submodule of the Noetherian vector space $M(\mathfrak{C})$. Then, for ρ sufficiently large, the maximum number of vectors of degree $\leq \rho$ of $M(\mathfrak{C})$ which are 1-linearly independent mod. M is given by a polynomial $a_0 \binom{\rho}{d} + a_1 \binom{\rho}{d-1} + \dots + a_d$ where a_0, \dots, a_d are whole numbers independent of ρ and where $a_0 > 0$. The $M(\mathfrak{C})$ -degree a_0 of M is the sum of the $M(\mathfrak{C})$ -degrees of the components $M(\mathfrak{C}) \cap Q_j$, where the Q_j 's are those primary components of the representation $M = M(\mathfrak{C}) \cap Q_1 \cap \dots \cap Q_h$ whose radicals are d -dimensional.*

If M is zero dimensional as a submodule of the vector space $M(\mathfrak{C})$, the polynomial of Theorem 2.34 reduces to a_0 . This gives the following theorem.

THEOREM 2.35. *If the arbitrary module $M \subseteq V$ is zero-dimensional as a submodule of the Noetherian vector space $M(\mathfrak{C})$, the factor module $M(\mathfrak{C})/M$ has finite P -rank a_0 . That is, $M(\mathfrak{C})/M$ is a Noetherian vector space over P as scalar domain which is generated by a_0 elements. This rank a_0 is the sum of the finite P -ranks of the factor modules $M(\mathfrak{C})/M(\mathfrak{C}) \cap Q_j$ where the $M(\mathfrak{C}) \cap Q_j$ are the primary components of M is a submodule of $M(\mathfrak{C})$.*

If M is a module of V , the $M(\mathfrak{C})$ -characteristic function and the $M(\mathfrak{C})$ -degree of M denote the $M_0(\mathfrak{C}_0)$ -characteristic function and $M_0(\mathfrak{C}_0)$ -degree

of M_0 . Hence, the $M(\mathbb{C})$ -characteristic function of a module $M \subseteq V$ is, for ρ sufficiently large, the maximum number of vectors of degree $\leq \rho$ of $M(\mathbb{C})$ which are P -linearly independent mod. M . The same terminology will be used for ideals of $P[x_1, \dots, x_n]$.

2.4. Connection between length and degree. Let M be any module of the vector space V and let M be \mathfrak{p} -primary as a submodule of the Noetherian vector space $M(\mathbb{C})$. (As in [1], \mathfrak{p} -primary means primary with \mathfrak{p} as radical.) We know from [1], Theorem 2.41, that there exists a \mathfrak{p} -primary composition sequence of length l , $M \subset M_1 \subset \dots \subset M_{l-1} \subset M(\mathbb{C})$. If \mathfrak{p} is zero-dimensional, \mathfrak{p} is maximal and hence, according to [1], Theorem 2.41, the factor module M_{i+1}/M_i is then operator isomorphic with $P[x_1, \dots, x_n]/\mathfrak{p}$ with respect to $P[x_1, \dots, x_n]$ as operator domain and hence certainly with respect to P as operator domain. The factor module $P[x_1, \dots, x_n]/\mathfrak{p}$ is an extension field of P of finite rank a_0^* , where this rank is the degree, in the sense of the Hilbert characteristic function, of \mathfrak{p} as a sub-ideal of $P[x_1, \dots, x_n]$. It follows immediately that the degree a_0 of M as a \mathfrak{p} -primary submodule of $M(\mathbb{C})$ where \mathfrak{p} is 0-dimensional satisfies $a_0 = la_0^*$. It is the purpose of this section to extend this statement to the case when \mathfrak{p} is d -dimensional (for ideals, see [8] Sections 29 through 32).

Let t be a new variable and let $V(t)$ denote the m -dimensional row vector space over $P(t)[x_1, \dots, x_n]$ as scalar domain. If M is a module of V , $M(t)$ denotes the module $V(t)$ which is generated by the vectors of M .

LEMMA 2.41. *Let M be a \mathfrak{p} -primary module as a submodule of the Noetherian vector space $M(\mathbb{C})$. Let ξ_1, \dots, ξ_n be a general point of \mathfrak{p} where ξ_1 is transcendental with respect to P . Then, $M^* = (M(t), (x_1 - t)M(t)(\mathbb{C}))$ is $(\mathfrak{p}(t), x_1 - t)$ -primary as a submodule of $M^*(\mathbb{C})$. Furthermore, the length of M as a submodule of $M(\mathbb{C})$ is equal to the length of M^* as a submodule of $M^*(\mathbb{C})$.*

Proof. As in [8], Section 30, we assert that $V \circ M^* = M$. All we have to show is that $V \circ M^* \subseteq M$. However, if $v \in V \circ M^*$, then $\phi(t)v = \sum_j \pi_j(t)m_j + (x_1 - t)w$ where $\phi(t)$, $\pi_j(t)$ and the components of the vector $w \in V(t)$ are polynomials of $P[t]$ and where $m_j \in M$. Hence, for $t = x_1$ we conclude that $\phi(x_1)v = \sum_j \pi_j(x_1)m_j \in M$. Since ξ_1 is transcendental with respect to P , $\phi(x_1) \notin \mathfrak{p}$ and hence $v \in M$ which proves the assertion. In the same way, the remainder of the proof of Lemma 2.41 is an immediate extension of [8], Section 30, to modules.

For H -modules of F , Lemma 2.41 can be formulated as follows. Let

$F(t)$ denote the m -dimensional row vector space over $P(t)[x_0, \dots, x_n]$ as scalar domain. If M is an H -module of F , $M(t)$ denotes the H -module of $F(t)$ which is generated by the vectors of M . Since $M(t)$ is then generated by the vector forms of M , $M(t)$ is certainly homogeneous.

LEMMA 2.42. *Let M be an H -module of F which is \mathfrak{p} -primary as a submodule of $M(\mathbb{C})$. Let $\lambda, \lambda\xi_1, \dots, \lambda\xi_n$ be a general point of the H -ideal \mathfrak{p} where ξ_1 is transcendent with respect to P . Let $M^* = (M(t), (x_1 - tx_0)M(t)(\mathbb{C}))$ and $\mathfrak{p}^* = (\mathfrak{p}(t), x_1 - tx_0)$. Then, M^*_0 is \mathfrak{p}^*_0 -primary as a submodule of $M^*_0(\mathbb{C})$. Furthermore, the length of M as a submodule of $M(\mathbb{C})$ is the same as the length of M^*_0 as a submodule of $M^*_0(\mathbb{C})$.*

Proof. The proof is an immediate extension of [8], Section 31, to modules.

We need one more lemma to obtain the above stated purpose of this section. Let K be an extension field of P . Then V_K denotes the m -dimensional row vector space over $K[x_1, \dots, x_n]$ as scalar domain and F_K denotes the m -dimensional row vector space over $K[x_0, \dots, x_n]$ as scalar domain. If $M \subseteq V$, M_K denotes the module of V_K which is generated by the vectors of M and the same notation is used for H -modules of F .

LEMMA 2.43. *Let M be an H -module of F . Then, the $M(\mathbb{C})$ -characteristic function of M is equal to the $M_K(\mathbb{C})$ -characteristic function of M_K . The same holds if M is a module of V .*

Proof. The proof is an immediate extension of the proof of Theorem 23 in [7].

THEOREM 2.41. *Let M be an H -module of F which is \mathfrak{p} -primary as a submodule of $M(\mathbb{C})$. Let the $M(\mathbb{C})$ -degree of M be a_0 , the $M(\mathbb{C})$ -length of M be l , and the degree of \mathfrak{p} as an H -ideal of $P[x_0, \dots, x_n]$, a^*_0 . Then, $a_0 = la^*_0$. The same holds if M is a module of V .*

Proof. The proof is an immediate extension of [8], Section 32, to modules.

2.5. Systems of linear equations. In [1], Definition 2.51, the \mathfrak{p} -elementary divisor $e(\mathfrak{p})$, the \mathfrak{p} -exponent $\rho(\mathfrak{p})$ and the \mathfrak{p} -length $l(\mathfrak{p})$ of a module $M \subseteq V$ were defined in terms of the modules $M(\mathfrak{p})$ and $M'(\mathfrak{p})$. We now define the \mathfrak{p} -degree $a_0(\mathfrak{p})$ in the same way.

DEFINITION 2.51. *Let M be a module of V ; let \mathfrak{p} be a prime ideal of*

$P[x_1, \dots, x_n]$; and let $M(\mathfrak{p})$ and $M'(\mathfrak{p})$ be defined as in [1]. Then, the \mathfrak{p} -degree $a_0(\mathfrak{p})$ of M is the degree of $M(\mathfrak{p})$ as a submodule of $M'(\mathfrak{p})$.

Since $M(\mathfrak{p})$ is \mathfrak{p} -primary as a submodule of $M'(\mathfrak{p})$ (see [1] Lemma 2.51), $a_0(\mathfrak{p}) = l(\mathfrak{p})a^*_0$ where a^*_0 is the degree of \mathfrak{p} as an ideal of $P[x_1, \dots, x_n]$. Hence, for polynomial modules, the criterion of lengths (see [1], Section 2.6) can be stated as follows.

THEOREM 2.52. *Let $M_1 \subseteq M_2$ be two modules of V . Then, $M_1 = M_2$ if and only if M_1 and M_2 have the same associated primes and, for each associated prime \mathfrak{p} , the same \mathfrak{p} -degree.*

The zero module of the vector space V is a prime module and hence the developments of [1], Section 2.7, can be used. Since the radical of the zero module is the zero ideal of $P[x_1, \dots, x_n]$, Theorem 2.71 of [1] becomes the following.

THEOREM 2.53. *Let $M_1 \subseteq M_2$ be two modules of V . Then, $M_1 = M_2$ if and only if M_1 and M_2 have the same rank, the same non-zero associated primes, and, for each such associated prime $\mathfrak{p} \neq 0$, the same \mathfrak{p} -degree.*

Let $A = (\alpha_{ij})$ be an $m \times s$ matrix, where $\alpha_{ij} \in P[x_1, \dots, x_n]$ and $i = 1, \dots, m$ and $j = 1, \dots, s$. The columns of A generate a module C , called the column space of A , of the m -dimensional column vector space over $P[x_1, \dots, x_n]$ as scalar domain. The associated primes, \mathfrak{p} -length, \mathfrak{p} -elementary divisor, \mathfrak{p} -exponent, \mathfrak{p} -degree and the rank of A are defined as those of C . The criterion of solvability and Theorem 2.72 of [1] then become the following.

THEOREM 2.54. *Let $\sum_{j=1}^s \alpha_{ij}z_j = \gamma_i$, where $i = 1, \dots, m$ and α_{ij} and γ_i are polynomials of $P[x_1, \dots, x_n]$, be a system of m linear equations for the s unknowns z_j . Then this system has a solution $z_j \in P[x_1, \dots, x_n]$ ($j = 1, \dots, s$) if and only if the $m \times s$ matrix $A = (\alpha_{ij})$ and the augmented $m \times (s+1)$ matrix $B = (\alpha_{ij}, \gamma_i)$ have the same associated primes and, for each associated prime \mathfrak{p} , the same \mathfrak{p} -degree. This is equivalent to saying that A and B must have the same rank, the same non-zero associated primes and, for each non-zero associated prime \mathfrak{p} , the same \mathfrak{p} -degree.*

The notion of the transformed module M^0 of a module $M \subseteq V$ is discussed in [2]. The resultant of M^0 is investigated in Sections 5 and 6 of [2]. We shall use the term "norm" instead of "resultant" since this polynomial will be shown in Lemma 2.51 to be the complete analogue of the norm for

the case $n = 1$ (see [1], Definition 1.13). If we call the norm of M^0 also the norm of M , Theorem 2.53 of [2] states that the system of linear equations can be solved if and only if A and B have the same rank and norm. The following lemma shows that this criterion of [2] and the present Theorem 2.54 are equivalent.

LEMMA 2.51. *The non-zero associated primes $\mathfrak{p}_1, \dots, \mathfrak{p}_h$ of a module $M \subseteq V$ are uniquely determined by the irreducible factors π_1, \dots, π_h of the norm $\Delta = \prod_{j=1}^h \pi_j^{r_j}$ of M . The polynomial π_j is the elementary divisor, in the sense of [2] Section 5, of \mathfrak{p}_j . The degree of the primary factor $\pi_j^{r_j}$ of Δ is the \mathfrak{p} -degree $a_0(\mathfrak{p})$ of M . If the characteristic p of P is zero, the degree of π_j is the degree a^*_0 of \mathfrak{p}_j and the exponent l_j is then the \mathfrak{p}_j -length $l(\mathfrak{p}_j)$ of M .*

Proof. Let $\prod_{j=1}^h \pi_j^{r_j}$ be the factorization of the norm Δ of M . For the case of ideals, that is when the dimension of V is one, it was proved in [9], Sections 3 and 5, that there exists a one to one correspondence between the non-zero associated primes $\mathfrak{p}_1, \dots, \mathfrak{p}_h$ of the ideal and the irreducible factors π_1, \dots, π_h of Δ , where π_j is the elementary divisor of \mathfrak{p}_j . These proofs can be extended to modules without difficulty. From [9], Theorem XIII, it then follows that, if $p = 0$, the degree of π_j is the degree of the transform of \mathfrak{p}_j and hence is equal to the degree a^*_0 of \mathfrak{p}_j . We know that, for any characteristic $p = 0$ or $p \neq 0$, the \mathfrak{p}_j -degree $a_0(\mathfrak{p}_j)$ of M satisfies $a_0(\mathfrak{p}_j) = a^*_0 l(\mathfrak{p}_j)$ where $l(\mathfrak{p}_j)$ is the \mathfrak{p}_j -length of M . Hence, all that remains to be proved is that, for arbitrary characteristic of P , the degree of $\pi_j^{r_j}$ is $a_0(\mathfrak{p}_j)$. For this purpose, we consider the $n + 1$ closures $\text{Cl}_{i-1}(M^0)$, $i = 1, \dots, n + 1$, of the transformed module M^0 which are defined in [2], Section 3. If M is an arbitrary module of V , the intersection of the primary components of dimensions $n - 1, n - 2, \dots, n - i + 1$ of a Noether decomposition of M is an isolated component of M which is denoted by $\text{Cl}_{i-1}(M)$. It can be easily shown that then $(\text{Cl}_{i-1}(M))^0 = \text{Cl}_{i-1}(M^0)$ where M^0 is the transform of M and $(\text{Cl}_{i-1}(M))^0$ the transform of $\text{Cl}_{i-1}(M)$ (see [9], Section 5, for the proof of the corresponding theorem on ideals which can be immediately extended to modules). We assert that, if Q is a \mathfrak{p} -primary component of M where \mathfrak{p} has dimension $n - i$, then the \mathfrak{p} -degree $a_0(\mathfrak{p})$ of M is equal to the degree a'_0 of $Q' = \text{Cl}_{i-1}(M) \cap Q$ as a submodule of $\text{Cl}_{i-1}(M)$ (we omit the lower index j in \mathfrak{p}_j). Since Q' is \mathfrak{p} -primary as a submodule of $\text{Cl}_{i-1}(M)$, we know that $a'_0 = l' a^*_0$ where l' is the length of Q' as a submodule of $\text{Cl}_{i-1}(M)$ and a^*_0 is the degree of \mathfrak{p} . Hence, all we have to show is that $l' = l(\mathfrak{p})$ where $l(\mathfrak{p})$ is the length of $M(\mathfrak{p})$ as a submodule of $M'(\mathfrak{p})$. However,

$Q' = M(\mathfrak{p}) \circ Q_1 \circ \cdots \circ Q_s$ where Q_1, \cdots, Q_s are those primary components of a Noether decomposition of M whose radicals $\mathfrak{p}_1, \cdots, \mathfrak{p}_s$ have dimensions larger than $n-i$ and where $\mathfrak{p}_z \not\subseteq \mathfrak{p}$ for $1 \leq z \leq s$. In the same way, $\text{Cl}_{i-1}(M) = M'(\mathfrak{p}) \circ Q_1 \circ \cdots \circ Q_s$. Now, let $M(\mathfrak{p}) \subset M_1 \subset \cdots \subset M_{l-1} \subset M'(\mathfrak{p})$ be a \mathfrak{p} -primary composition sequence from $M(\mathfrak{p})$ to $M'(\mathfrak{p})$ of length $l = l(\mathfrak{p})$. Then, if $L = Q_1 \circ \cdots \circ Q_s$, $M(\mathfrak{p}) \circ L \subset M_1 \circ L \subset \cdots \subset M_{l-1} \circ L \subset M'(\mathfrak{p}) \circ L$ can be proved immediately to be a composition sequence from Q' to $\text{Cl}_{i-1}(M)$ which proves the assertion. According to [2], M^0 arises from M by adjunction of new variables and an isomorphism from which we conclude that the degree of $\text{Cl}_{i-1}(M^0) \circ Q^0 = Q'^0$ as a submodule of $\text{Cl}_{i-1}(M^0)$ is also $a_0(\mathfrak{p})$ (the upper index 0 denotes as before the transformed module).

Let $y_j = \sum_{h=1}^n u_{jh} x_h$ be the transformed variables where $1 \leq j \leq n$ and where the $u^2 u_{jh}$ are new variables which are adjoined to P (see [2], Section 2). The module Q^0 is \mathfrak{p}^0 -primary as a submodule of $\text{Cl}_{i-1}(M^0) \circ Q^0$ and, if η_1, \cdots, η_n is a general point of \mathfrak{p}^0 , then $\eta_{i+1}, \cdots, \eta_n$ are algebraically independent with respect to $P(u)$ and each η_1, \cdots, η_i is algebraic with respect to $P_\eta = P(u, \eta_{i+1}, \cdots, \eta_n)$ (see [9], Section 3). Every module M^0 gives rise to a module M_η of the m -dimensional row vector space T_η over $P_\eta[y_1, \cdots, y_i]$ as scalar domain. Here, M_η is generated by the vectors of M^0 after the substitution $y_h = \eta_h$ for $i+1 \leq h \leq n$ has been carried out. Then, $(\text{Cl}_{i-1}(M^0))_\eta \circ Q^0_\eta = Q'^0_\eta$ and Q'^0_η is \mathfrak{p}^0_η -primary as a submodule of $(\text{Cl}_{i-1}(M^0))_\eta$ where \mathfrak{p}^0_η is the zero-dimensional prime ideal of $P_\eta[y_1, \cdots, y_i]$ whose general point is η_1, \cdots, η_i (see [6], Section 97, where the proofs are carried out for ideals). If $Q^0 \subset H^0_1 \subset \cdots \subset H^0_{l-1} \subset \text{Cl}_{i-1}(M^0)$ is a \mathfrak{p}^0 -primary composition sequence from Q^0 to $\text{Cl}_{i-1}(M^0)$, $Q'^0_\eta \subset (H^0_1)_\eta \subset \cdots \subset (H^0_{l-1})_\eta \subset (\text{Cl}_{i-1}(M^0))_\eta$ can easily be shown to be a \mathfrak{p}^0_η -primary composition sequence from Q'^0_η to $(\text{Cl}_{i-1}(M^0))_\eta$ and hence the length of Q'^0_η as a submodule of $(\text{Cl}_{i-1}(M^0))_\eta$ is still $l(\mathfrak{p})$. Since \mathfrak{p}^0_η is zero-dimensional, the factor module $P_\eta[y_1, \cdots, y_i]/\mathfrak{p}^0_\eta$ has finite P_η -rank; and since \mathfrak{p}^0 is transformed, this rank is equal to the degree of \mathfrak{p}^0 (this statement is false for not-transformed ideals). Consequently, the degree of Q'^0_η as a submodule of $(\text{Cl}_{i-1}(M^0))_\eta$ is still $a_0(\mathfrak{p})$. Since Q'^0_η is zero dimensional as a submodule of $(\text{Cl}_{i-1}(M^0))_\eta$, we conclude that $a_0(\mathfrak{p})$ is the finite P_η -rank of the factor module $(\text{Cl}_{i-1}(M^0))_\eta/Q'^0_\eta$. In [9], Sections 3 and 5, it was proved for the case of ideals that the degree of the primary factor $\pi_j{}^{\tau_j}$ of Δ is the P_η -rank of $(\text{Cl}_{i-1}(M^0))_\eta/Q'^0_\eta$, which proof carries over to modules. Hence, Lemma 2.51 is proved.

Remark 2.51. According to Lemma 2.51, $\tau_j = l(\mathfrak{p}_j)$ only if $p = 0$. However we know that, if $n = 1$, always $\tau_j = l(\mathfrak{p}_j)$ for any characteristic

(see [1], Definition 1.13). The reason for this is that the dimension $n - i$ of \mathfrak{p}_i is zero if $n = 1$ and hence $i = 1$. Theorem XIV of [9] implies that, if $i = 1$, Theorem XIII of [9] still holds and hence $\tau_j = l(\mathfrak{p}_j)$. Consequently, we can say that, if \mathfrak{p}_j is an $(n - 1)$ -dimensional associated prime of M , $\tau_j = l(\mathfrak{p}_j)$ for any characteristic.

Let us call the norm of the column space of a matrix the norm of the matrix. Then, to say that A and B of Theorem 2.54 have the same rank and norm is the same as saying that A and B have the same rank, non-zero associated primes and \mathfrak{p} -degrees, which shows the equivalence of Theorem 5.3 of [2] and the present Theorem 2.54. Therefore, Theorem 5.3 of [2] is a consequence of the "criterion of lengths" of [1].

2.6. The elementary divisor, the \mathfrak{p} -exponent and the exponents of partial differentiation. The notation is the same as in the proof of Lemma 2.51. The \mathfrak{p} -elementary divisor $e(\mathfrak{p})$ of M is defined in [1], Section 2.5. The elementary divisor of M (without \mathfrak{p}) denotes the elementary divisor of M^0 in the sense of Section 5 of [2]. The elementary divisor E of M has the same irreducible factors as the resultant, i.e., $E = \Pi_{j=1}^h \pi_j^{\sigma_j}$ (see [2], Section 5).

LEMMA 2.61. *The primary factor $\pi_i^{\sigma_i}$ of E which corresponds to the associated prime \mathfrak{p} of M is the elementary divisor of the \mathfrak{p} -elementary divisor $e(\mathfrak{p})$ of M . If the characteristic of P is zero, σ_i is the \mathfrak{p} -exponent $\rho(\mathfrak{p})$ of M .*

Proof. The primary factor $\pi_i^{\sigma_i}$ is the highest common factor of the polynomials contained in $P(u)[y_1, \dots, y_n] \cap (Q^0 : \text{Cl}_{i-1}(M^0))$ (see [9], Sections 1 and 5, where the proofs are carried out for ideals). However, $Q^0 = M^0(\mathfrak{p}) \cap L^0$ and $\text{Cl}_{i-1}(M^0) = M'^0(\mathfrak{p}) \cap L^0$ and hence $Q^0 : \text{Cl}_{i-1}(M^0) = (M^0(\mathfrak{p}) \cap L^0) : (M'^0(\mathfrak{p}) \cap L^0) = (M^0(\mathfrak{p}) : (M'^0(\mathfrak{p}) \cap L^0)) \cap (L^0 : (M'^0(\mathfrak{p}) \cap L^0)) = M^0(\mathfrak{p}) : (M'^0(\mathfrak{p}) \cap L^0)$. From the fact that the associated primes of L are not contained in \mathfrak{p} , it follows that $M^0(\mathfrak{p}) : M'^0(\mathfrak{p}) \cap L^0 = M^0(\mathfrak{p}) : M'^0(\mathfrak{p}) = e^0(\mathfrak{p})$. Hence, $\pi_i^{\sigma_i}$ is the highest common factor of the polynomials contained in $P(u)[y_1, \dots, y_n] \cap e^0(\mathfrak{p})$ which is the elementary divisor of $e^0(\mathfrak{p})$, i.e. of $e(\mathfrak{p})$ (see [9], Section 1). Since $e(\mathfrak{p})$ is \mathfrak{p} -primary and since $\rho(\mathfrak{p})$ is defined as the ordinary exponent of $e(\mathfrak{p})$, it follows from Theorem 7.1 of [3] that $\sigma_i = \rho(\mathfrak{p})$ in case P has characteristic zero.

The \mathfrak{p} -modal rank $\lambda(\mathfrak{p})$ of M and the \mathfrak{p} -modal exponent $\delta(\mathfrak{p})$ of M are essential for the theory of partial differential equations (see Part I).

LEMMA 2.62. *If the characteristic p of P is zero, the \mathfrak{p} -modal exponent $\delta(\mathfrak{p})$ is the smallest integer such that $M'(\mathfrak{p}) \cap \mathfrak{p}^{\delta(\mathfrak{p})} \subseteq M(\mathfrak{p})$ and, if*

furthermore \mathfrak{p} is an isolated associated prime of M , $\delta(\mathfrak{p}) = \rho(\mathfrak{p})$. The \mathfrak{p} -modal rank $\lambda(\mathfrak{p})$ of M is, after a suitable linear transformation and for any characteristic $p = 0$ or $\neq 0$, equal to the \mathfrak{p} -degree $a_0(\mathfrak{p})$ of M .

Proof. Using the notation of 1.2, the ideals $\mathfrak{p}^{**}(\xi^{(1)}), \dots, \mathfrak{p}^{**}(\xi^{(e)})$ are conjugated ideals with respect to the relative automorphisms of P^{**} with respect to P . Since the modules $M'^{**}(\mathfrak{p})$, $M^{**}(\mathfrak{p})$ and V^{**} are invariant under these relative automorphisms, $\delta(\mathfrak{p})$ could also have been defined in Definition 1.22 as the smallest integer such that

$$M'^{**}(\mathfrak{p}) \cap \left(\bigcap_{j=1}^e \mathfrak{p}^{**}(\xi^{(j)}) \right)^{\delta(\mathfrak{p})} V^{**} \subseteq M^{**}(\mathfrak{p}).$$

Since the ideals $\mathfrak{p}^{**}(\xi^{(j)})$ are relatively prime,

$$\bigcap_{j=1}^e \mathfrak{p}^{**}(\xi^{(j)}) = \Pi_{j=1}^e \mathfrak{p}^{**}(\xi^{(j)});$$

and, since $p = 0$, $\Pi_{j=1}^e \mathfrak{p}^{**}(\xi^{(j)}) = \mathfrak{p}^{**}$ (see [7], Theorem 25). Hence, $\delta(\mathfrak{p})$ is the smallest integer such that $M'^{**}(\mathfrak{p}) \cap (\mathfrak{p}^{**})^{\delta(\mathfrak{p})} V^{**} \subseteq M^{**}(\mathfrak{p})$. Since the asterisks denote an algebraic extension of \bar{P} , it can easily be proved that they may be omitted for the definition of $\delta(\mathfrak{p})$. If \mathfrak{p} is an isolated associated prime of M , $M'(\mathfrak{p}) = V$ and hence $\delta(\mathfrak{p})$ is the smallest integer such that $\mathfrak{p}^{\delta(\mathfrak{p})} M'(\mathfrak{p}) \subseteq M(\mathfrak{p})$, which means that $\delta(\mathfrak{p})$ is then the ordinary exponent of $\mathfrak{e}(\mathfrak{p})$, i. e. $\delta(\mathfrak{p}) = \rho(\mathfrak{p})$. The \mathfrak{p} -modal rank $\lambda(\mathfrak{p})$ is the P^{**} -rank of $M'^{**}(\mathfrak{p})/M^{**}(\mathfrak{p})$ and hence, according to Lemma 2.43, the \bar{P} -rank of $\overline{M'(\mathfrak{p})}/\overline{M(\mathfrak{p})}$. This rank is, according to Theorem 2.41, equal to la_0 , where l is the length of $\overline{M(\mathfrak{p})}$ as a submodule of $\overline{M'(\mathfrak{p})}$ and a_0 is the \bar{P} -rank of $\bar{P}[x_1, \dots, x_i]/\bar{\mathfrak{p}}$. Since $l = l(\mathfrak{p})$ and since, after a suitable linear transformation, a_0 is the degree of \mathfrak{p} , $\lambda(\mathfrak{p}) = a_0(\mathfrak{p})$ after a suitable linear transformation, which completes the proof.

2.7. General remarks. In the foregoing theory we have investigated those properties of a matrix $A = (\alpha_{ij})$, where $\alpha_{ij} \in P[x_1, \dots, x_n]$, which are determined by the associated primes, the $M(\mathfrak{p})$, $M'(\mathfrak{p})$, $\mathfrak{e}(\mathfrak{p})$, $l(\mathfrak{p})$, $\rho(\mathfrak{p})$ and $a_0(\mathfrak{p})$, of the row space and column space of A . We shall refer to these ideals and integers as the invariants of A . If $n = 1$, these invariants can be computed from the sub-determinants of A (see [1]) and hence must be the same for the column space and the row space. The following example shows that the column space and row space of A may already have different associated primes when $n = 2$ and hence that the sub-determinants of A can not be used to compute the invariants of A .

Example 2.71. Consider the 2×2 matrix $A = (\alpha_{ij})$, where $\alpha_{ij} \in P[x, y]$ and where $\alpha_{11} = x$, $\alpha_{12} = 0$, $\alpha_{21} = y$, $\alpha_{22} = 0$. The column space is generated by the one column vector (x, y) and hence is a closed module which has only one associated prime, namely the zero ideal (ω) of $P[x, y]$. The row space M is generated by the row vectors $(x, 0)$ and $(y, 0)$. Hence $\text{Cl}(M)$ is generated by the row vector $(1, 0)$ and it can be easily shown that M has two associated primes, namely $\mathfrak{p}_1 = (\omega)$ and $\mathfrak{p}_2 = (x, y)$.

It can be easily shown that, if $A = (\alpha_{ij})$ is an $m \times s$ matrix, $\alpha_{ij} \in P[x_1, \dots, x_n]$, the invariants of A are invariant under similarity transformations PAQ where P and Q are square, invertible polynomial matrices. If $n = 1$, a Noether decomposition of the row space (or column space) of A gives rise to invariants, namely the classical elementary divisors of different rank, which determine A completely to within similarity transformation. Whether this is still the case for $n > 1$ is an unsolved problem. Equally unsolved, for $n > 1$, is the question whether invariants which arise from a Noether decomposition of a module $M \subseteq V$ determine the factor module V/M to within an isomorphism.

UNIVERSITY OF SOUTHERN CALIFORNIA,
LOS ANGELES, CALIFORNIA.

REFERENCES

-
- [1] E. Snapper, "Polynomial matrices in one variable, differential equations and module theory," *American Journal of Mathematics*, vol. 69 (1947), pp. 299-326.
 - [2] E. Snapper, "The resultant of a linear set," *American Journal of Mathematics*, vol. 66 (1944), pp. 59-68.
 - [3] E. Snapper, "Partial differentiation and elementary divisors," *Annals of Mathematics*, vol. 46 (1945), pp. 388-408.
 - [4] R. A. Frazier, W. J. Duncan and A. R. Collar, *Elementary Matrices*, Cambridge University Press (1938).
 - [5] P. M. Grundy, "A generalization of additive ideal theory," *Proceedings of the Cambridge Philosophical Society*, vol. 38 (1941), pp. 241-279.
 - [6] B. L. v. d. Waerden, *Moderne Algebra*, vol. 2, Berlin, Springer, 1940.
 - [7] B. L. v. d. Waerden, "Eine Verallgemeinerung des Bézoutschen Theorems," *Mathematische Annalen*, vol. 99 (1928), pp. 497-541.
 - [8] B. L. v. d. Waerden, "On Hilbert's function, series of composition of ideals and a generalisation of the theorem of Bézout," *Proceedings, Akademie van Wetenschappen te Amsterdam*, vol. 31 (1928), pp. 749-770.
 - [9] E. Noether, "Eliminationstheorie und allgemeine Idealtheorie," *Mathematische Annalen*, vol. 90 (1923), pp. 229-261.

PROJECTIVITIES OF FINITE PROJECTIVE PLANES.*

By REINHOLD BAER.

The group of projectivities of a finite projective plane Π has been investigated fairly thoroughly under the hypothesis that Π is the projective plane over some Galois Field.¹ But once this hypothesis is dropped, the machinery of analytic geometry is not available any more; and comparatively little² is known concerning the projectivities of Π .

In the present investigation we undertake to study projectivities of finite projective planes Π without any further hypotheses on the nature of Π . Consequently we shall not be able to make use of such devices as Galois Fields and their automorphisms, the calculus of matrices and so on. In their stead we shall utilize the configuration of the fixed elements of a projectivity and the arithmetical properties of the invariant n of Π —this number n is determined by the fact that every line in Π carries $1 + n$ points and every point is on $1 + n$ lines; it seems to play the same rôle in the theory of finite projective planes as is played by the order in the theory of finite groups.

The system of the fixed elements of a projectivity is closed under the operations of joining and intersecting; and it contains as many points as it contains lines. This last fact is not at all trivial, since its analogue for groups of projectivities fails to be true. The structural properties of the system of fixed elements provide us with a useful principle for classifying projectivities (Section 1).

Every projectivity effects a permutation in the set of all the points of Π and at the same time in the set of all the points of any given line which it leaves invariant. This makes it possible to apply, in a variety of ways, certain congruences relating orders and characters of permutations. These—together with some applications—we have collected in Section 2.

With these tools we are able to evolve (in Section 3) a fairly complete

* Received January 7, 1946; Revised March 25, 1946. Presented to the American Mathematical Society.

¹ See, for instance, Carmichael (1), Chapter XII.

² Some such material may be found in Baer (2). The work of Steck (1) is narrower in its scope than ours, since he tacitly presupposes the validity of the Theorem of Desargues, though he neither states nor uses this hypothesis, even in the proofs of some theorems that fail to be true without such a hypothesis.

theory of the projectivities of prime power order. Their simplest and most striking property is the fact that the number of their fixed points is never two. If the order is a power of the prime p , then $p \leq 1 + n + n^2$ and $1 + n < p$ imply the absence of fixed elements whereas $p = 2$ implies their existence. If the p^i -th power of a projectivity of order a power of p does not leave invariant complete subplanes of Π , then p^{i+1} is a divisor of $(1 + n + n^2)n^2(n^2 - 1)$.

In Section 4 we apply these results to groups Δ of projectivities with the property that the identity is the only projectivity in Δ which leaves invariant a complete subplane of Π . [The best known instance of such a group is the so-called group of special collineations³ of a projective plane over a Galois Field.] We show that the order of such a group Δ is always a divisor of $(1 + n + n^2)(1 + n)n^3(1 - n)^2$; and if the order of Δ happens to be a divisor of one of the four factors of the above limit, then the nature of the projectivities in Δ may be determined.

Notations.

If Π is a projective plane,⁴ then we denote by $P + Q$ the line connecting the two different points P and Q ; and we denote by hk the point of intersection of the two different lines h and k .

A projectivity ϕ of Π is a 1 — 1 correspondence, mapping the point P in Π upon the point $P\phi$ in Π and the line h in Π upon the line $h\phi$ in Π and meeting the following requirement:

P is on h if, and only if, $P\phi$ is on $h\phi$.

We are considering finite projective planes. Each of the lines in such a finite projective plane Π carries $n + 1$ points and each point is on $n + 1$ lines. The total number of points (lines) in Π is $1 + n + n^2$. This integer n (≥ 2) retains its significance throughout.

Projectivities of finite projective planes are necessarily of finite order; and we denote by $o(\phi)$ the *order* of the projectivity ϕ . Since ϕ is completely determined by the permutation it induces in the set of the $1 + n + n^2$ points in Π , it follows that

$$o(\phi) \leq 1 + n + n^2.$$

³ For this group, see Carmichael (1), section 93, and Jacobson (1), p. 80.

⁴ For the concept of "projective plane" as used here, see Baer (1), (2); Carmichael (1), Chapter XI; Hall (1); Veblen-Young (1).

1. The structure of the system of fixed elements of a projectivity. A point or line left invariant by the projectivity ϕ of Π will be termed a fixed point or a fixed line of ϕ ; and we denote by $N(\phi)$ the total number of fixed points of ϕ .

THEOREM 1. $N(\phi)$ is the number of fixed lines of ϕ .

Proof. The point P and the line h are said to form a pair with respect to ϕ , if P is both on h and on $h\phi$; and we denote by $M(\phi)$ the number of distinct pairs with respect to ϕ . Denote furthermore by $N'(\phi)$ the number of fixed lines of ϕ .

If h is a fixed line of ϕ , then P, h is a pair if, and only if, P is on h . Thus $(n+1)N'(\phi)$ is the number of pairs P, h such that h is a fixed line of ϕ . If h is not a fixed line of ϕ , then h and $h\phi$ meet in a well determined point $h(h\phi)$; and $h(h\phi), h$ is the only pair containing h . Hence $1+n+n^2-N'(\phi)$ is the total number of pairs P, h such that h is not a fixed line. Combining these two results we find that

$$M(\phi) = (n+1)N'(\phi) + 1 + n + n^2 - N'(\phi) = nN'(\phi) + 1 + n + n^2.$$

The point P and the line h are said to form a couple with respect to ϕ , if h passes both through P and through $P\phi$; and we denote by $M'(\phi)$ the number of distinct couples with respect to ϕ . Since couple is the dual of pair, it follows by dualization that

$$M'(\phi) = nN(\phi) + 1 + n + n^2.$$

The point P and the line h form a pair with respect to ϕ if, and only if, they form a couple with respect to ϕ^{-1} . Hence

$$M(\phi) = M'(\phi^{-1}).$$

But ϕ and ϕ^{-1} have the same fixed elements, proving $N(\phi) = N(\phi^{-1})$. Thus we find finally that

$$\begin{aligned} nN(\phi) + 1 + n + n^2 &= nN(\phi^{-1}) + 1 + n + n^2 = M'(\phi^{-1}) \\ &= M(\phi) = nN'(\phi) + 1 + n + n^2; \end{aligned}$$

and this implies $N(\phi) = N'(\phi)$, as we desired to show.

We denote by $\Phi(\phi)$ the system of all the fixed elements of the projectivity ϕ . This system is closed under the operations of connecting points by a line and of intersecting lines. It contains as many points as it contains

lines [Theorem 1]. Consequently ϕ may belong to one and only one of the following four types.⁵

TYPE A. *The fixed element free projectivities.*

They are characterized by $N(\phi) = 0$.

TYPE B. *The generalized elations.*

There exists a fixed line h and a fixed point H on h such that every fixed point is on h , every fixed line passes through H . Either H and h are the only fixed elements; or else h is the only fixed line carrying more than one fixed point and H is the only fixed point which is on more than one fixed line. The line h and the point H are, therefore, uniquely determined by the projectivity; and may, consequently, be called its *axis* and *center* respectively.

TYPE C. *The generalized homologies.*

It will be convenient to distinguish two possibilities.⁶

TYPE C'. $N(\phi) \neq 3$.

There exists a fixed line h and a fixed point H , not on h , of ϕ with the following properties: every fixed point, not H , is on h ; every fixed line, not h , passes through H . The fixed line h carries $N(\phi) - 1 \neq 2$ fixed points; and every fixed line, different from h , (if any), carries exactly two fixed points. The fixed point H is on $N(\phi) - 1 \neq 2$ fixed lines; and every fixed point, different from H , (if any) is on exactly two fixed lines. Thus H and h are uniquely determined by ϕ ; and may therefore be termed the *center* and *axis* of ϕ respectively.

Remark 1. It may happen that ϕ is a projectivity of Type B or of Type C' and that at the same time $N(\phi) = 1$.

Remark 2. Suppose that $N(\phi) = 2$. Then ϕ possesses two fixed points U and V and two fixed lines u and v ; and we may select notations in such a way that $u = U + V$ and $U = uv$. Then ϕ is a projectivity of Type B with axis u and center U ; and at the same time ϕ is of Type C' with axis v and center V . If, conversely, the projectivity ϕ is at the same time of Type B and of Type C', then one verifies readily that $N(\phi) = 2$. This ambiguity

⁵ For an enumeration of the possible types of "degenerate projective planes," see e. g. Hall (1), p. 232. Steck (1) points out a possibility of classifying projectivities according to the number of fixed points.

⁶ Sometimes another subdivision of Type C will be found more convenient; see 3, Theorem 3 below.

in notation will not create any confusion, mainly because of 3, Theorem 4, (2) below.

TYPE C''. $N(\phi) = 3$.

Then the system of fixed elements of ϕ is just an ordinary (not degenerate) triangle.⁷

TYPE D. *The projectivities whose system of fixed elements is a projective subplane of Π .*

It is well known⁸ that the projectivity ϕ is of Type D if, and only if, there exist four fixed points of ϕ no three of which are collinear. If ϕ is of Type D, then each of its fixed lines carries $i + 1$ fixed points; and each of its fixed points is on $i + 1$ fixed lines. Furthermore

$$N(\phi) = 1 + i + i^2 \text{ and } i = \frac{1}{2}(-1 + (4N(\phi) - 3)^{\frac{1}{2}});$$

and we shall say that ϕ is of Type (D, i) .

A projectivity ϕ of Type D is completely determined by the system $\Phi(\phi)$ of its fixed elements and by the permutation it induces in the set of points on some fixed line. If ϕ is of Type (D, i) , then it has $i + 1$ fixed points on every fixed line; and thus it follows that

$$o(\phi) \leq n - i \leq n - 2, \text{ if } \phi \text{ is of Type } (D, i).$$

Remark 3. If the projectivity ϕ belongs to Types C or D, then a proof of Theorem 1 may be effected by a more direct use of the structural properties of the configuration $\Phi(\phi)$. But such a procedure would break down, if ϕ belongs to Types A or B.

Remark 4. One may be tempted to prove an analogue of Theorem 1 for groups of projectivities. If there exist three non-collinear points that are left invariant by all the projectivities in the group Δ , then it is easy to show that Δ has as many fixed points as it has fixed lines (Types C and D). But without this hypothesis the analogue of Theorem 1 fails to be true. For consider as examples the projective plane over a Galois Field and the following groups of projectivities of this plane:

1. All the projectivities which leave invariant a given point P ; one fixed point, no fixed line.

⁷ I. e., the three fixed points are not collinear.

⁸ Baer (2), Hall (1).

2. All the projectivities which leave invariant two different points P and Q ; two fixed points, one fixed line.

3. All the projectivities which leave invariant every point on a given line h ; $n + 1$ fixed points, one fixed line.

THEOREM 2. *If ϕ^k is of Type B or C', then the axis and the center of ϕ^k are fixed elements of ϕ .*

Proof. If x is a fixed element of ϕ^k , then $(x\phi)\phi^k = (x\phi^k)\phi = x\phi$, proving that ϕ maps fixed elements of ϕ^k upon fixed elements of ϕ^k . Thus ϕ induces a projectivity in $\Phi(\phi^k)$. But axis and center of ϕ^k are uniquely determined by projective properties of the configuration $\Phi(\phi^k)$. Hence they are fixed elements of ϕ .

COROLLARY 1. *If ϕ^k is of Type A or B or C', then ϕ is of Type A or B or C respectively.*

Proof. The system $\Phi(\phi)$ of the fixed elements of ϕ is part of the system $\Phi(\phi^k)$ of the fixed elements of ϕ^k . If ϕ^k is of Type A, then it is fixed element free, implying that ϕ is fixed element free. If ϕ^k is of Type B or C', then it possesses an axis h and a center H ; and it follows from Theorem 2 that h and H are fixed elements of ϕ too. Now one deduces our contention easily from the definitions of the types.

COROLLARY 2. $N(\phi) = N(\phi^k)$ whenever $N(\phi^k) < 3$.

Proof. This is obvious if $N(\phi^k) = 0$ (see Corollary 1). It is a consequence of Theorem 2 in case $N(\phi^k) = 1$, since this implies that ϕ^k is of Type B or C'. If, finally, $N(\phi^k) = 2$, then one verifies (as in the proof of Theorem 2) that ϕ maps fixed elements of ϕ^k upon fixed elements of ϕ^k . But one of the two fixed points of ϕ^k is on two fixed lines of ϕ^k whereas the other one is on one and only one. Hence ϕ cannot interchange the two fixed points of ϕ^k , proving that they are fixed points of ϕ . Thus $N(\phi^k) = 2$ implies $N(\phi) = 2$. (See Remark 2.)

COROLLARY 3. *Suppose that $N(\phi^k) = 3$. Then*

- (a) $N(\phi) \neq 2$.
- (b) $N(\phi) = 1$ implies $N(\phi^2) = 3$ and $k = 2k'$ for some k' .
- (c) $N(\phi) = 0$ implies $N(\phi^3) = 3$ and $k = 3k''$ for some k'' .
- (d) *If ϕ^k is of Type C'' and if $o(\phi) \div o(\phi^k)$ is prime to 6, then ϕ is of Type C''.*

Proof. As in the proof of Theorem 2 we show that ϕ induces a permutation of the three fixed points of ϕ^k . The order t of this permutation is 1, 2 or 3 and correspondingly $N(\phi)$ is 3, 1 or 0, since every fixed point of ϕ is a fixed point of ϕ^k . This completes the proofs of (a), (b) and (c), if one remembers that ϕ^k is necessarily a power of ϕ^t and that t is a divisor of $o(\phi)$. To prove (d) we have only to note that $N(\phi) = 3$, if the hypotheses of (d) are satisfied and that therefore ϕ and ϕ^k have the same fixed elements.

COROLLARY 1. *If ϕ^k is of Type C and if $o(\phi) \div o(\phi^k)$ is prime to 3, then ϕ is of Type C.*

Proof. This is a consequence of Corollary 1 in case ϕ^k is of Type C'. Thus we may assume that ϕ^k is of Type C'' and possesses, therefore, exactly three fixed points P, Q, R which are not collinear. It is a consequence of our hypothesis and Corollary 3, (c) that ϕ possesses fixed points too; and thus, necessarily, one of the three points P, Q, R , say P , is left invariant by ϕ . As to Q and R , they are either both left invariant by ϕ or else they are interchanged by ϕ . In either case the line $R + Q$ is a fixed line of ϕ . Since the fixed point P of ϕ is not on the fixed line $Q + R$ of ϕ , it is shown again that ϕ is of Type C.

2. Characters of permutations. If γ is a permutation of the finite set S , then it is customary⁹ to term the number of fixed elements of γ the character $\chi(\gamma)$ of γ . In this section we shall deduce some formulas on these characters which we shall need for several applications in the future. It may be said here already that the set S may either be the set of all the points of a projective plane or the set of all the points on a fixed line of a projectivity.

THEOREM 1.¹⁰ *If γ is a permutation of a finite set S and p a prime number, then*

$$\chi(\gamma^{p^i}) \equiv \chi(\gamma^{p^j}) \text{ modulo } p^{i+1} \text{ for } 0 \leq i < j.$$

Proof. Every element x in S belongs to a cycle of the permutation γ . This cycle consists of p^i elements in S if, and only if, x is left invariant by γ^{p^i} , but not by $\gamma^{p^{i-1}}$. Since the cycles are mutually exclusive sets, the number of elements in S which are left invariant by γ^{p^i} , but not by $\gamma^{p^{i-1}}$, is a multiple of p^i ; and this fact may be restated as follows:

⁹ Speiser (1), p. 18.

¹⁰ This theorem is probably known to many people. But the author did not succeed in locating a convenient reference for it. Speiser (1), Satz 102 is a much deeper theorem; and the deduction of the present formula from it is not quite immediate either.

$$(1) \quad \chi(\gamma^{p^i}) \equiv \chi(\gamma^{p^{i-1}}) \text{ modulo } p^i.$$

If $i < j$, then we infer from (1) that

$$\chi(\gamma^{p^j}) \equiv \chi(\gamma^{p^{j-1}}) \text{ modulo } p^j$$

and this implies in particular that

$$\chi(\gamma^{p^{j-1}}) \equiv \chi(\gamma^{p^j}) \text{ modulo } p^{i+1}.$$

Consequently we have

$$\chi(\gamma^{p^i}) \equiv \chi(\gamma^{p^{i+1}}) \equiv \cdots \equiv \chi(\gamma^{p^j}) \text{ modulo } p^{i+1},$$

completing the proof.

COROLLARY 1. If γ is a permutation of the finite set S , p a prime number, $0 \leq i < p$ and $\chi(\gamma^{p^i}) < p^{i+1}$, then

$$\chi(\gamma^{p^i}) = \chi(\gamma^{p^j}).$$

Proof. Since every fixed element of γ^{p^i} is a fixed element of γ^{p^j} , we may deduce from our hypothesis that

$$0 \leq \chi(\gamma^{p^j}) - \chi(\gamma^{p^i}) \leq \chi(\gamma^{p^j}) < p^{i+1}.$$

On the other hand it follows from Theorem 1 that $\chi(\gamma^{p^j}) - \chi(\gamma^{p^i})$ is a multiple of p^{i+1} . Hence $\chi(\gamma^{p^j}) = \chi(\gamma^{p^i})$.

A great number of congruences with composite modulus may be derived from Theorem 1. To obtain a unified derivation of them, we consider the class of number theoretical functions¹¹ $f(n)$ with the following property:

(P) If k is a positive integer, if p is a prime number, and if $0 \leq i < j$, then $f(kp^i) \equiv f(kp^j) \text{ modulo } p^{i+1}$.

Noting that $\gamma^{kp^i} = (\gamma^k)^{p^i}$ we infer from Theorem 1 that the characters $\chi(\gamma^i) = t(i)$ of the powers of a given permutation are number-theoretical functions with Property (P). Another example is furnished—by Euler's Theorem—by the powers of a given integer. Finally it may be worth noting that the sum, difference and product of number theoretical functions with the property (P) are again number theoretical functions with the property (P) so that *these functions form a ring*.

In order to simplify the enunciation of our next theorem, we have to introduce a few symbols. If m is a positive integer, then we denote by m^* the product of all the different prime divisors of m . Thus m^* is the l. c. m.

¹¹ These are integral valued functions of an integral variable.

of all the squarefree divisors of m . Next we need a generalization of Möbius' function, namely the function $\nu_D(d)$ which we define for divisors d of the integer D only by the following rules:

$\nu_D(d) = 0$, if d and Dd^{-1} are not relatively prime;

$\nu_D(d) = 1$, if d and Dd^{-1} are relatively prime and if the number of different prime divisors of d is even;

$\nu_D(d) = -1$, if d and Dd^{-1} are relatively prime and if the number of different prime divisors of d is odd.

THEOREM 2. *If D is a divisor of the positive integer m and a multiple of m^* , and if the number theoretical function $f(n)$ meets the requirement (P), then*

$$\sum_{d|D} \nu_D(d) f(md^{-1}) \equiv 0 \text{ modulo } mD^{-1}m^*.$$

Proof. Suppose that the prime number p is a divisor of m . Denote by p^h the highest power of p , dividing D ; and denote by p^{h+k} the highest power of p , dividing m . Then $0 < h$, since p is a divisor of m^* and, therefore, of the multiple D of m^* ; and $0 \leq k$, since D is a divisor of m .

(i) If the divisor d of D is prime to p , then $\nu_D(d) = -\nu_D(dp^h)$.

Since d and p are relatively prime, and since p^h is the highest power of p , dividing D , it follows that the g. c. d. of d and Dd^{-1} is the same as the g. c. d. of dp^h and $D(dp^h)^{-1}$. Hence $\nu_D(d) = 0$ if, and only if, $\nu_D(dp^h) = 0$. Since d is prime to p , and since h is positive, dp^h is divisible by just one prime number more than d ; and now (i) is an immediate consequence of the definition of ν_D .

(ii) If the divisor d of D is prime to p , then

$$\nu_D(d) f(md^{-1}) + \nu_D(dp^h) f(m(dp^h)^{-1}) \equiv 0 \text{ modulo } p^{k+1}.$$

Since d and p are prime, dp^{k+h} is a divisor of m ; and hence $j = md^{-1}p^{-k-h}$ is an integer. Now we deduce from (i) and Property (P) that

$$\begin{aligned} \nu_D(d) f(md^{-1}) + \nu_D(dp^h) f(m(dp^h)^{-1}) &\equiv \nu_D(d) [f(jp^{k+h}) - f(jp^k)] \\ &\equiv 0 \text{ modulo } p^{k+1}. \end{aligned}$$

(iii) $\sum_{d|D} \nu_D(d) f(md^{-1}) \equiv 0 \text{ modulo } p^{k+1}.$

Every divisor of D has the form dp^j where d is prime to p and $0 \leq j \leq h$. The divisors of D that are prime to p are exactly the divisors of $D' = Dp^{-h}$. Noting finally that $\nu_D(dp^j) = 0$ whenever d is a divisor of D' and $0 < j < h$, since in this case both dp^j and $D(dp^j)^{-1}$ are divisible by p , we find that

$$\sum_{d|D} v_D(d) f(md^{-1}) = \sum_{d|D'} [v_D(d) f(md^{-1}) + v_D(dp^h) f(m(dp^h)^{-1})];$$

and (iii) is an immediate consequence of (ii).

But p^{k+1} is the highest power of the prime number p which divides $mD^{-1}m^*$. Thus it follows from (iii) that every prime power divisor of $mD^{-1}m^*$ is a divisor of $\sum_{d|D} v_D(d) f(md^{-1})$. Hence $mD^{-1}m^*$ itself is a divisor, completing the proof of this theorem.

Remark 1. Note that $d=1$ is always a divisor of D and $v_D(1) = 1$. Thus $f(m)$ itself appears always as a term in the sum of Theorem 2.

Remark 2. We may let, in particular, $D = m$ in Theorem 2. Then the sum ranges over all the divisors of m and the modulus of the congruence is m^* .

The most interesting special case of Theorem 2 involves in its statement the use of Möbius' function $\mu(d)$. It should be remembered that $\mu(d) = 0$ if d is not squarefree, that $\mu(d) = 1$, if d is the product of an even number of different primes and that $\mu(d) = -1$, if d is the product of an odd number of different primes.

COROLLARY 2. *If the number-theoretical function $f(m)$ meets the requirement (P), then*

$$\sum_{d|m} \mu(d) f(md^{-1}) \equiv 0 \text{ modulo } m.$$

Proof. If d is a divisor of m , then $\mu(d) \neq 0$ is a necessary and sufficient condition for d to be a divisor of m^* . If d is a divisor of m^* , then d and m^*d^{-1} are relatively prime so that $v_{m^*}(d) = \mu(d) \neq 0$. From these remarks we infer that

$$\sum_{d|m} \mu(d) f(md^{-1}) = \sum_{d|m^*} v_{m^*}(d) f(md^{-1}).$$

Considering now the special case $D = m^*$ of Theorem 2, we find that this sum is divisible by $mD^{-1}m^* = m$.

COROLLARY 3. *If γ is a permutation of the finite set S , if $o(\gamma) = m$ is the order of the permutation γ and M the number of elements in S , and if $\chi(\gamma^{mp^{-1}}) < q$ for every prime divisor p of m and every prime divisor q of mp^{-1} , then $\chi(\gamma) \equiv M$ modulo m .*

Proof. If d is a divisor of m , $1 < d$, and if q is a prime divisor of md^{-1} , then it follows from our hypothesis that $\chi(\gamma^{md^{-1}}) < q$. Hence we infer from Corollary 1 that $\chi(\gamma^{md^{-1}}) = \chi(\gamma^{m(dq)^{-1}})$. Now it follows by complete induction that $\chi(\gamma) = \chi(\gamma^{mi^{-1}})$ whenever $i \neq 1$ is a divisor of m . Remembering that

$\chi(\gamma^m) = \chi(1) = M$ and that $\sum_{d|m} \mu(d) = 0$ for $m \neq 1$, we deduce now from Corollary 2 (and Theorem 1) that $\chi(\gamma) \equiv M$ modulo m .

If ϕ is a projectivity of the projective plane Π , then ϕ effects a permutation of the $1 + n + n^2$ points in Π ; and the number $N(\phi)$ of the fixed points of ϕ is the character of this permutation. It is now quite clear how to use the results of the present section for obtaining fixed point formulas for projectivities. The following application, however, does not seem to be without interest.

If the projectivity ϕ is of Type D, then there exists a uniquely determined integer $i(\phi)$ such that every fixed line of ϕ carries $i(\phi) + 1$ fixed points of ϕ . Furthermore every power of ϕ is of Type D too so that the numbers $i(\phi^k)$ are well determined.

THEOREM 3. *If the projectivity ϕ is of Type D, then*

$$\sum_{d|o(\phi)} \mu(d) i(\phi^{o(\phi)d^{-1}}) \equiv 0 \text{ modulo } o(\phi).$$

Proof. Consider a fixed line h of ϕ . Then ϕ effects a permutation of the $n + 1$ points of h ; and it is a consequence of Theorem 1 that the number-theoretical function $i(\phi^k) + 1$ meets the requirement (P). Hence $i(\phi^k)$ meets the requirement (P) too; and now our Theorem is a consequence of Corollary 2.

3. Projectivities of order a power of a prime. Apart from the number $N(\phi)$ of the fixed points (lines) of the projectivity ϕ we shall make use of the number $N(\phi, h)$ of all the fixed points of ϕ which are situated on the line h . The discussion of this section will be based on the following proposition.

THEOREM 1. *If the projectivity ϕ of the finite projective plane Π is of order $o(\phi) = p^m$ for p a prime, and if $0 \leq i < m$, then*

- (a) $N(\phi^{p^i}) \equiv 1 + n + n^2 \text{ modulo } p^{i+1}$; and
- (b) $N(\phi^{p^i}, h) \equiv 1 + n \text{ modulo } p^{i+1}$ for every fixed line h of ϕ .

Proof. ϕ^{p^i} effects a permutation of the $1 + n + n^2$ points of Π whose character is $N(\phi^{p^i})$. Hence it follows from 2, Theorem 1, that $N(\phi^{p^i}) \equiv N(\phi^{p^m}) \equiv 1 + n + n^2 \text{ modulo } p^{i+1}$, proving (a). If h is a fixed line of ϕ , then ϕ and its powers induce permutations of the $n + 1$ points on h ; and the characters of these permutations are the numbers $N(\phi^j, h)$. Hence it follows from 2, Theorem 1, that $N(\phi^{p^i}, h) \equiv N(\phi^{p^m}, h) \equiv n + 1 \text{ modulo } p^{i+1}$, proving (b).

COROLLARY 1. *If the projectivity ϕ is of prime power order $o(\phi) = p^m$, and if $n + 1 < o(\phi)$, then*

- (a) $N(\phi) = 0$ or 1; and
- (b) $N(\phi) = 1$ if, and only if, $\phi^{p^{m-1}}$ is a perspectivity; and
- (c) $m = 1$ implies $N(\phi) = 0$.

Proof. If h is a fixed line of ϕ , then ϕ effects a permutation of the $n + 1$ points on h . Thus it follows from 2, Corollary 1, (or Theorem 1, (b)) that $N(\phi^{p^{m-1}}, h) = n + 1$. Hence every point on the fixed line h of ϕ is a fixed point of $\phi^{p^{m-1}}$. But it is impossible that a projectivity, not 1, leaves invariant all the points on two different lines. Hence $N(\phi) \leq 1$, proving (a). If $N(\phi) = 1$, then the points on the fixed line of ϕ are fixed points of $\phi^{p^{m-1}}$; and it is well known¹² that a projectivity, not 1, which leaves invariant all the points on a certain line is a perspectivity. Thus $\phi^{p^{m-1}}$ is a perspectivity whenever $N(\phi) = 1$. If, conversely, $\phi^{p^{m-1}}$ is a perspectivity, then it follows from 1, Theorem 2, that its axis and center are fixed elements of ϕ so that $N(\phi) \neq 0$; and this implies $N(\phi) = 1$ by (a), completing the proof of (b). Since perspectivities leave invariant more than 1 point, $m = 1$ and $N(\phi) = 1$ are incompatible by (b); and (c) is a consequence of (a).

THEOREM 2A. *If $o(\phi) = p^m$ for p a prime, and if ϕ^{p^i} is fixed element free, then $1 + n + n^2 \equiv 0$ modulo p^{i+1} .*

This is a consequence of the fact that by Theorem 1, (a)

$$0 \equiv N(\phi^{p^i}) \equiv 1 + n + n^2 \text{ modulo } p^{i+1}.$$

COROLLARY 2. *Projectivities of order a power of 2 possess fixed elements.*

For in this case we may infer from Theorem 1, (a) that $N(\phi)$ is odd, since $1 + n + n^2$ is always odd.

COROLLARY 3 *If $1 + n + n^2$ is a prime, then $o(\phi) = 1 + n + n^2$ is a necessary and sufficient condition for ϕ to be a fixed element free projectivity of prime power order.*

Proof. If $N(\phi) = 0$ and $o(\phi) = p^m$ for p a prime, then it follows from Theorem 2A that p is a prime divisor of the prime number $1 + n + n^2$ so that $p = 1 + n + n^2$. Hence ϕ is a permutation of p points whose order is $p^m \neq 1$, proving that $m = 1$ and $o(\phi) = 1 + n + n^2$. If, conversely, ϕ is

¹² E. g. Baer (1), p. 140, Corollary 2.3.

- of order the prime number $p = 1 + n + n^2$, then it effects a cyclic permutation of the p points in Π , implying $N(\phi) = 0$.

Remark 1. If Π happens to be the projective plane over a Galois Field $GF(n)$, then there exists¹³ a projectivity γ which effects a cyclic permutation of the $1 + n + n^2$ points in Π . Clearly γ and all its powers are fixed element free, showing the impossibility of improving Theorem 2A. The order of γ is the greatest possible order of projectivities of Π , since every projectivity effects a permutation of the $1 + n + n^2$ points whose order certainly cannot exceed $1 + n + n^2$.

THEOREM 2B. *If $o(\phi) = p^m$ for p a prime and ϕ^{p^i} is of Type B, then $n^2 \equiv 0$ modulo p^{i+1} ; and $n \equiv 0$ modulo p^{i+1} , if either $N(\phi) \neq 1$ or $N(\phi^{p^i}) = 1$.*

Proof. Since ϕ^{p^i} is of Type B, all its fixed points are on its axis h and all its fixed lines pass through its center H . It follows from 1, Theorem 2, that H and h are fixed elements of ϕ too. Now we deduce from Theorem 1 that

$$1 + n + n^2 \equiv N(\phi^{p^i}) \equiv N(\phi^{p^i}, h) \equiv n + 1 \text{ modulo } p^{i+1}.$$

Consequently $n^2 \equiv 0$ modulo p^{i+1} . If furthermore $N(\phi^{p^i}) = 1$, then $n \equiv 0$ modulo p^{i+1} may be deduced from the above congruences.

Assume, finally, that $N(\phi) \neq 1$. Since h is a fixed line of ϕ , there exists a fixed line w of ϕ such that $w \neq h$. But every fixed element of ϕ is a fixed element of ϕ^{p^i} ; and H is consequently the one and only fixed point of ϕ^{p^i} which is on w . Thus it follows from Theorem 1, (b) that

$$1 \equiv N(\phi^{p^i}, w) \equiv n + 1 \text{ modulo } p^{i+1}$$

or $n \equiv 0$ modulo p^{i+1} , completing the proof.

Remark 2. It is easy to construct projectivities of prime power order $p^m = n^2$ whose p^{m-1} -st power is of Type B (in which case p^m is not a divisor of n).

COROLLARY 4. *Suppose that ϕ is a projectivity of prime power order and that n is a prime number. Then ϕ is of Type B if, and only if, $o(\phi) = n$ or n^2 .*

Proof. If $o(\phi) = p^m$ for p a prime number, and if ϕ is of Type B, then it follows from Theorem 2B that p is a divisor of n . But n is a prime

¹³ Singer (1), p. 379.

number, proving $n = p$. If h is a fixed line of ϕ and $0 \leq i < m$, then it follows from Theorem 1, (b) that

$$1 \equiv 1 + n \equiv N(\phi^{p^i}, h) \text{ modulo } p.$$

But the only numbers between 0 and $n + 1$ that are congruent to 1 modulo $p = n$ are 1 and $n + 1$, proving that $N(\phi^{p^i}, h)$ is either 1 or $n + 1$. Since h is a fixed line of ϕ and, therefore, of ϕ^{p^i} , this implies that ϕ^{p^i} is of Type B too. Hence it follows from Theorem 2B that p^m is a divisor of $n^2 = p^2$ so that $o(\phi)$ is either n or n^2 .

Assume, conversely, that $o(\phi) = n$. Then it follows from Theorem 1, (a) that $N(\phi) \equiv 1 + n + n^2 \equiv 1$ modulo n so that, in particular, $N(\phi) \neq 0$. If h is a fixed line of ϕ , then it follows from Theorem 1, (b) that $N(\phi, h) \equiv n + 1 \equiv 1$ modulo n ; and it follows as before that $N(\phi, h)$ is either 1 or $n + 1$. But a projectivity possessing fixed elements whose fixed lines carry 1 or $n + 1$ fixed points, is a projectivity of Type B.

If, finally, $o(\phi) = n^2$, then $o(\phi^n) = n$; and it follows from the result of the preceding paragraph of the present proof that ϕ^n is of Type B. Hence it is a consequence of 1, Corollary 1, that ϕ is of Type B.

If we had at the same time $o(\phi) = n^2$ and $N(\phi) \neq 1$, then there would exist a fixed line k of ϕ , not the axis of ϕ (nor of ϕ^n). Thus the center of ϕ and ϕ^n would be the only fixed point on k ; and we could infer from Theorem 1, (b) that

$$1 \equiv N(\phi^n, k) \equiv n + 1 \text{ modulo } n^2$$

which is impossible. Thus we have not only completed the proof of Corollary 4, but also of the following proposition.

COROLLARY 4'. *If n is a prime number, and if $o(\phi) = n$, then ϕ is either an elation (in the strict sense) or else $N(\phi) = 1$; and $o(\phi) = n^2$ implies $N(\phi) = 1$.*

THEOREM 2C'. *Suppose that ϕ is of prime power order $o(\phi) = p^m$, and that ϕ^{p^i} is of Type C'.*

- (a) *If $N(\phi) \neq 1$, then $n \equiv 1$ modulo p^{i+1} .*
- (b) *If p is odd and $N(\phi) = 1$, then $n \equiv -1$ modulo p^{i+1} and $N(\phi^{p^i}) = 1$.*
- (c) *If $p = 2$ and $N(\phi^{2^i}) = 1$, then $n \equiv -1$ modulo 2^{i+1} .*
- (d) *If $p = 2$ and $N(\phi) = 1 \neq N(\phi^{2^i})$, then either $N(\phi^{2^i}) \neq 1$ and $n \equiv 1$ modulo 2^i or else $N(\phi^{2^{i-1}}) = 1$ and $n \equiv -1$ modulo 2^i .*

Proof. Since ϕ^{p^i} is of Type C', it possesses an axis h and a center H ,

not on h , which are by 1, Theorem 2, fixed elements of ϕ . All the fixed points of ϕ^{p^i} with the exception of H are on h ; and all the fixed lines of ϕ^{p^i} , not h , pass through H . Hence it follows from Theorem 1 that

$$1 + n + n^2 \equiv N(\phi^{p^i}) \equiv N(\phi^{p^i}, h) + 1 \equiv (n + 1) + 1 \text{ modulo } p^{i+1} \text{ or}$$

$$(e) \quad n^2 \equiv 1 \text{ modulo } p^{i+1}.$$

If $N(\phi) \neq 1$, then there exists a fixed line w of ϕ which is different from the axis h of ϕ^{p^i} , since h is a fixed line of ϕ . This line w carries two and only two fixed points of ϕ^{p^i} , namely H and the intersection wh . Hence it follows from Theorem 1, (b) that

$$2 \equiv N(\phi^{p^i}, w) \equiv n + 1 \text{ modulo } p^{i+1} \text{ or } n \equiv 1 \text{ modulo } p^{i+1},$$

proving (a).

Suppose finally that $N(\phi) = 1$. Then we infer from (e) and Theorem 1, (a) that

$$1 \equiv N(\phi) \equiv 1 + n + n^2 \equiv 2 + n \text{ modulo } p \text{ or } n \equiv -1 \text{ modulo } p.$$

If p is odd, then this implies that p and $n - 1$ are relatively prime. From (e) we infer that $(n + 1)(n - 1) \equiv n^2 - 1 \equiv 0 \text{ modulo } p^{i+1}$; and this implies $n \equiv -1 \text{ modulo } p^{i+1}$, proving the first part of (b). If $N(\phi^{p^i})$ were different from 1, then we could apply (a) on $\phi^{p^i} = (\phi^{p^i})^{p^0}$; and we would obtain $n \equiv 1 \text{ modulo } p$ which is incompatible with $n \equiv -1 \text{ modulo } p^{i+1}$, since $p \neq 2$. Hence $N(\phi^{p^i}) = 1$, completing the proof of (b).

If $p = 2$ and $N(\phi^{2^i}) = 1$, then the fixed line h of ϕ does not carry any fixed points of ϕ^{2^i} ; and we infer from Corollary 1, (b) that

$$0 \equiv N(\phi^{2^i}, h) \equiv n + 1 \text{ modulo } 2^{i+1},$$

proving (c).

If finally $p = 2$ and $N(\phi^{2^i}) \neq 1 = N(\phi)$, then there exists a positive integer $k \leq i$ such that $N(\phi^{2^k}) \neq 1 = N(\phi^{2^{k-1}})$. Applying (a) on ϕ^{2^k} and $(\phi^{2^k})^{2^{i-k}} = \phi^{2^i}$, we find that $n \equiv 1 \text{ modulo } 2^{i-k+1}$; and applying (c) on ϕ and $\phi^{2^{k-1}}$ we find that $n \equiv -1 \text{ modulo } 2^k$. Since it is impossible that n is modulo 4 congruent to $+1$ as well as to -1 , it follows that $k = 1$ or $i = k$. If $k = 1$, then $N(\phi^2) \neq 1$; and applying (a) upon ϕ^2 and $(\phi^2)^{2^{i-1}} = \phi^{2^i}$ we see that $n \equiv 1 \text{ modulo } 2^i$. If $i = k$, then $N(\phi^{2^{i-1}}) = 1$; and it follows from (c) that $n \equiv -1 \text{ modulo } 2^i$, completing the proof.

THEOREM 2C''. Suppose that ϕ is of prime power order $o(\phi) = p^m$, and that ϕ^{p^i} is of Type C''.

- (a) If $p \neq 2, 3$, then $n \equiv 1$ modulo p^{i+1} and $N(\phi) = 3$.
 (b) If $p = 2$, then $n \equiv 1$ modulo 2^{i+1} and $N(\phi) = 1$ or 3 .
 (c) If $p = 3$, then $N(\phi) \neq 0$ implies $n \equiv 1$ modulo 3^{i+1} and $N(\phi) = 3$;
 and $N(\phi) = 0$ implies $n \equiv 1$ modulo 3^i and $N(\phi^3) = 3$.

Proof. The system $\Phi(\phi^{p^i})$ is an ordinary, non degenerate triangle, since ϕ^{p^i} is of Type C''; and the system of fixed elements of ϕ^{p^j} for $0 \leq j \leq i$ is part of this triangle. We show first:

- (d) $N(\phi) \neq 0$ implies $n \equiv 1$ modulo p^{i+1} .

For if w is a fixed line of ϕ , then w carries two, and only two, fixed points of ϕ^{p^i} . Hence it follows from Theorem 1, (b) that $2 \equiv N(\phi^{p^i}, w) \equiv n + 1$ modulo p^{i+1} , proving (d).

If $p \neq 2, 3$, then we infer from 1, Corollary 3 that $N(\phi) = 3$; and (a) is a consequence of (d). If $p = 2$, then it follows from 1, Corollary 3 that $N(\phi) = 1$ or 3 ; and (b) is a consequence of (d). If finally $p = 3$, then we infer from 1, Corollary 3 that $N(\phi) = 0$ or 3 and that $N(\phi) = 0$ implies $N(\phi^3) = 3$. Now (c) is obtained by applying (d) on ϕ^3 , if $N(\phi) = 0$, and by applying (d) on ϕ , if $N(\phi) \neq 0$.

COROLLARY 5. If $n - 1$ is a prime number, then the projectivity ϕ of odd prime power order $o(\phi) = p^m$ has the following properties.

- (a) If ϕ is of Type C and $N(\phi) \neq 1$, then $o(\phi) = n - 1$.
 (b) If $o(\phi) = n - 1$ and $n \neq 4$, then ϕ is of Type C and $N(\phi) \neq 1$.

Proof. Suppose first that ϕ is of Type C and that $N(\phi) \neq 1$. Then it follows from Theorem 2C', (a) and Theorem 2C'' that $n \equiv 1$ modulo p . Thus p is a divisor of the prime number $n - 1$, proving $p = n - 1$. Consider a fixed line h of ϕ . Then it follows from Theorem 1, (b) that

$$N(\phi^{p^i}, h) \equiv n + 1 \equiv (n - 1) + 2 \equiv p + 2 \equiv 2 \text{ modulo } p.$$

Thus $N(\phi^{p^i}, h)$ is a number between 0 and $n + 1$ which is congruent to 2 modulo $p = n - 1$. Hence $N(\phi^{p^i}, h)$ is either 2 or $n + 1$, for every fixed line h of ϕ . Noting $1 < N(\phi) \leq N(\phi^{p^i})$ and noting the fact that the fixed points of ϕ , and therefore those of ϕ^{p^i} , are not all collinear, it follows that $\phi^{p^{m-1}}$ is of Type C. Hence it follows from $N(\phi) \neq 1$ and Theorem 2C', (a) and Theorem 2C'' that

$$n \equiv 1 \text{ modulo } p^m$$

so that $n - 1 = p$ is divisible by p^m . Hence $m = 1$, proving that $o(\phi) = p = n - 1$.

Assume, conversely, that $o(\phi) = n - 1 \neq 3$. Since $n - 1 = p$ is a prime number, it follows from Theorem 1, (a) that

$$N(\phi) \equiv 1 + n + n^2 \equiv 3 \text{ modulo } p.$$

If $N(\phi)$ were 0, then this would imply $n - 1 = p = 3$; and we excluded this possibility in our hypothesis. Hence $N(\phi) \neq 0$ and there exist fixed lines of ϕ . If h is a fixed line of ϕ , then it follows from Theorem 1, (b) that

$$N(\phi, h) \equiv n + 1 \equiv 2 \text{ modulo } p.$$

But $N(\phi, h)$ is a number between 0 and $n + 1 = p + 2$, proving that $N(\phi, h)$ is either 2 or $n + 1$. Consequently $N(\phi) \neq 1$ and ϕ is of Type C, completing the proof.

Remark 3. If Π is the projective plane over the Galois Field $\text{GF}(4)$, then $n = 4$. If r is a number neither 0 nor 1 in $\text{GF}(4)$, then the projectivity ϕ mapping the point with coordinates (x_0, x_1, x_2) upon the point with coordinates (rx_2, x_0, x_1) is of order $3 = n - 1$ and is fixed point free, showing that the hypothesis $n \neq 4$ cannot be omitted in (b). Note that it follows from the proof that we could have substituted the hypothesis $N(\phi) \neq 0$ for the hypothesis $n \neq 4$ in part (b).

COROLLARY 6. *If $n + 1$ is a prime number, then the following properties of the projectivity ϕ of prime power order $o(\phi) = p^m$ imply each other.*

(a) $o(\phi) = n + 1$.

(b) ϕ is of Type C and $N(\phi) = 1$.

Proof. Assume first that ϕ is of Type C and that $N(\phi) = 1$. Then it follows from Theorem 2C', (b) and (c) that $n \equiv -1 \text{ modulo } p$. Thus p is a divisor of the prime number $n + 1$ and hence $p = n + 1$. This implies in particular $p \neq 2$, since $2 \leq n$. If h is the fixed line of ϕ , then it follows from Theorem 1, (b) that

$$N(\phi^{p^i}, h) \equiv n + 1 \equiv p \equiv 0 \text{ modulo } p$$

so that $N(\phi^{p^i}, h)$ is either 0 or $n + 1$. Since $N(\phi, h) = 0$, and since ϕ possesses a fixed point not on h , it would follow from $N(\phi^{p^i}, h) = n + 1$ that $i \neq 0$ and that ϕ^{p^i} is a homology (Type C'). If $N(\phi^{p^i}, h) = 0$, then it follows from the definitions of the types that ϕ^{p^i} is of Type C' and $N(\phi^{p^i}) = 1$. In either case we deduce from Theorem 2C', (b) that $n \equiv -1 \text{ modulo } p^m$ so that p^m is a divisor of $n + 1 = p$. Hence $m = 1$ and $o(\phi) = p = n + 1$, proving that (a) is a consequence of (b).

Assume, conversely, that $o(\phi) = n + 1 = p$. Then p is odd; and we infer from Theorem 1, (a) that

$$N(\phi) \equiv 1 + n + n^2 \equiv 1 \text{ modulo } p$$

so that in particular $N(\phi) \neq 0$. If h is a fixed line of ϕ , then it follows from Theorem 1, (b) that

$$N(\phi, h) \equiv n + 1 \equiv \text{modulo } p$$

so that $N(\phi, h)$ is either 0 or $n + 1$. But if $N(\phi, h)$ were $n + 1$, then ϕ would be a perspectivity and there would exist fixed lines of ϕ carrying 1 or 2 fixed points, contradicting the fact, just proven, that every fixed line of ϕ carries 0 or $n + 1$ fixed points. Thus $N(\phi, h) = 0$ for every fixed line of ϕ , proving that $N(\phi) = 1$ and that ϕ is of Type C. Hence (b) is a consequence of (a).

Remark 4. Every power of a prime may be the integer n of a projective plane; but it is not known at present whether any other integer may be the n of a projective plane. If n is a prime power and $n - 1$ a prime, then either $n = 3$ or $n = 2^t$; and if n is a prime power and $n + 1$ a prime, then $n = 2^r$. We note that $n - 1$ is a prime, if $n = 3, 4, 8, 32, 128$ and that $n + 1$ is a prime for $n = 4, 16, 256$.

Remark 5. Denote by Z the group of projectivities of the projective plane Π which is generated by the perspectivities (in the strict sense of the word). This group has sometimes been termed ¹⁴ *the group of special collineations*. If Π is in particular the projective plane over the Galois Field $GF(n)$ consisting of exactly n elements, then it is well known ¹⁵ that the order of Z is $(1 + n + n^2)(1 + n)n^3(1 - n)^2$ and that the only projectivity of Type D in Z is the identity. If the prime number p is a divisor of the order Z , then Z contains an element of the order p . Thus it follows from the Corollaries 5 and 6 that the results obtained in Theorems 2C' and 2C'' are "best" results. The method employed by Singer (1) may be used to prove the existence of a projectivity of order $n + 1$ in Z .

In the following proposition we give a summary of some of the preceding results which seems of particular interest in the light of the preceding Remark 5.

COROLLARY 7. *If ϕ is of prime power order $o(\phi) = p^m$, and if ϕ^{p^t} is not of Type D, then,*

¹⁴ Jacobson (1), p. 80.

¹⁵ Carmichael (1), p. 358.

- (a) p^{i+1} is a divisor of $n^2(1+n+n^2)(n-1)(n+1)$; and
 (b) $N(\phi) \equiv 0$ or 1 or 3 modulo p^{i+1} .

Proof. If ϕ^{p^i} is of Types A or B, then we infer from Theorems 2A and 2B that p^{i+1} is a divisor of $1+n+n^2$ or n^2 respectively. If ϕ^{p^i} is of Type C', then we deduce from the statement (e), derived in the course of the proof of Theorem 2C' that p^{i+1} is a divisor of n^2-1 . If finally ϕ^{p^i} is of Type C'', then it follows from Theorem 2C'', (a) and (b) that p^{i+1} is a divisor of $n-1$, unless $p=3$ and $N(\phi)=0$. In the latter case 3^i is a divisor of $n-1$ and 3 is a divisor of $1+n+n^2$ so that 3^{i+1} is a divisor of $(n-1)(1+n+n^2)$, completing the proof of (a). (b) is readily deduced from the Theorems 2, since the exceptional cases of these theorems always involve that $N(\phi)=0$ or 1 or 3.

The extent to which the geometrical properties of a projectivity of order a power of p are determined by the arithmetical properties of the integers n and p will be made strikingly clear by the following theorem. Important applications will be made in the next section.

THEOREM 3. *Suppose that ϕ is of prime power order $o(\phi)=p^m$, and that ϕ^{p^i} , for $i < m$, is not of Type D.*

- (A) *If $p \neq 3$, then the following properties imply each other.*
 (A, 1) ϕ^{p^i} is of Type A;
 (A, 2) $1+n+n^2 \equiv 0$ modulo p^{i+1} ;
 (A, 3) $1+n+n^2 \equiv 0$ modulo p ;
 (A, 4) ϕ is of Type A.
 (B) *The following properties imply each other.*
 (B, 1) ϕ^{p^i} is of Type B;
 (B, 2) $n^2 \equiv 0$ modulo p^{i+1} ;
 (B, 3) $n \equiv 0$ modulo p ;
 (B, 4) ϕ is of Type B.
 (C*) *If p is odd, then the following properties imply each other.*
 (C*, 1) ϕ^{p^i} is of Type C and $N(\phi^{p^i})=1$;
 (C*, 2) $n \equiv -1$ modulo p^{i+1} ;
 (C*, 3) $n \equiv -1$ modulo p ;
 (C*, 4) ϕ is of Type C and $N(\phi)=1$.
 (C) *If $p \neq 2, 3$, then the following properties imply each other.*
 (C, 1) ϕ^{p^i} is of Type C and $N(\phi^{p^i}) \neq 1$;
 (C, 2) $n \equiv 1$ modulo p^{i+1} ;
 (C, 3) $n \equiv 1$ modulo p ;
 (C, 4) ϕ is of Type C and $N(\phi) \neq 1$.

Proof. It is a consequence of Theorem 2A that $(A, 1)$ implies $(A, 2)$; and it is obvious that $(A, 2)$ implies $(A, 3)$. It is a consequence of Theorem 2B that $(B, 1)$ implies $(B, 2)$; and it is obvious that $(B, 2)$ implies $(B, 3)$, since p is a prime number. It is a consequence of Theorem 2C', (b) and the oddness of p that $(C^*, 1)$ implies $(C^*, 2)$; and it is obvious that $(C^*, 2)$ implies $(C^*, 3)$. If $p \neq 2, 3$, then it follows from Theorem 2C', (a) and from Theorem 2C'', (a) that $(C, 1)$ implies $(C, 2)$; and it is obvious that $(C, 2)$ implies $(C, 3)$. If, finally, $p = 2$ or 3 , then we have $n^2 \equiv 1$ modulo p in the last two cases.

We shall make use of these implications during the remainder of the present proof.

If $p \neq 3$ and if $(A, 3)$ is valid, then $(B, 3)$, $(C^*, 3)$ and $(C, 3)$ cannot hold, since they would imply that $1 + n + n^2$ is congruent to 1 or 3 modulo p . Thus it follows from the implications already verified that ϕ^{p^t} cannot be of Types B or C. Hence ϕ^{p^t} is of Type A, proving that $(A, 1)$ is a consequence of $(A, 3)$ and $p \neq 3$. But $(A, 1)$ obviously implies $(A, 4)$; and that $(A, 4)$ implies $(A, 3)$, is one of the implications proved in the first paragraph of this proof, completing the proof of (A).

If $(B, 3)$ is satisfied by n , then $(A, 3)$, $(C^*, 3)$ and $(C, 3)$ are not satisfied. Thus ϕ^{p^t} cannot be of Types A or C. Hence it is of Type B, proving that $(B, 1)$ is a consequence of $(B, 3)$. By the same argument we see that ϕ is of Type B, proving that $(B, 4)$ is a consequence of $(B, 3)$. But $(B, 4)$ implies $(B, 3)$ by the results of the first paragraph of this proof; and $(B, 3)$ implies $(B, 1)$, completing the proof of (B).

If p is odd, and if $(C^*, 3)$ is satisfied by n , then neither $(A, 3)$ nor $(B, 3)$ nor $(C, 3)$ are satisfied by n . Thus ϕ and ϕ^{p^t} cannot be of Types A or B; and if they are of Type C, then $N(\phi) = N(\phi^{p^t}) = 1$. Hence $(C^*, 4)$ and $(C^*, 1)$ are consequences of $(C^*, 3)$; and it is clear how to complete the proof of (C^*) .

If, finally, $p \neq 2, 3$, and if $(C, 3)$ is satisfied by n , then $1 + n + n^2 \equiv 3$ modulo p so that neither $(A, 3)$ nor $(B, 3)$ nor $(C^*, 3)$ is satisfied by n . Thus ϕ and ϕ^{p^t} cannot be of Types A or B; and if they are of Type C, then $N(\phi)$ and $N(\phi^{p^t})$ are both different from 1. Consequently $(C, 3)$ implies $(C, 1)$ and $(C, 4)$; and it is clear how to complete the proof of (C).

Remark 6. If Π is the projective plane over the Galois Field $GF(q^k)$ for q a prime, then there exists a projectivity of Type D and order k . It is now easy to construct examples showing the indispensability of the hypothesis that ϕ^{p^t} be not of Type D.

A. Let q and k be primes such that $1 + 2q \equiv 0$ modulo k . Then $n = q^k$ and $1 + n + n^2$ is divisible by k , though there exists a projectivity of Type D and order k .

B. Let $q = k = p$, a prime. Then $n = p^p$ is divisible by p , though there exists a projectivity of Type D and order p .

C*. Let q and k be primes such that $q \equiv -1$ modulo k (e. g. $q = 19$ and $k = 5$). Then $n = q^k \equiv -1$ modulo k , though there exist projectivities of Type D and order k .

C. Let q and k be primes such that $q \equiv 1$ modulo k (e. g. $q = 11$ and $k = 5$). Then $n = q^k \equiv 1$ modulo k , though there exist projectivities of Type D and order k .

The projectivities possessing less than seven fixed points are of special interest. Note that some of their properties have been given in 1, Corollaries 2 and 3. The projectivities without fixed points are just those of Type A; and the projectivities of Type C, possessing 1 or 3 fixed points have found special treatment in Theorems 2C' and 2C''.

THEOREM 4. Suppose that ϕ is a projectivity of prime power order $o(\phi) = p^m$.

- (1) If ϕ is of Type B or C, then $N(\phi) \equiv 1, 3$ modulo p .
- (2) $N(\phi) \neq 2$.
- (3) If $N(\phi) = 3$, and if ϕ is of Type B, then $p = 2$.
- (4) If $N(\phi) = 4$, then $p = 3$ and ϕ is of Type B.
- (5) If $N(\phi) = 5$, then $p = 2$.
- (6) If $N(\phi) = 6$, then either ϕ is of Type B and $p = 5$ or else ϕ is of Type C and $p = 3$.

Proof. It is a consequence of Theorem 1, (a) that $N(\phi) \equiv 1 + n + n^2$ modulo p ; and (1) is an immediate consequence of Theorems 2B, 2C' and 2C''.

If $0 < N(\phi) < 7$, then ϕ is of Types B or C, since every projective plane contains at least 7 points. Thus we may make use of the statement (1) during the remainder of this proof.

(2) is a consequence of the fact that $2 \not\equiv 1, 3$ modulo p .

If $N(\phi) = 3$ and ϕ is of Type B, then we deduce from Theorem 1, (a) and Theorem 2B that

$$3 \equiv 1 + n + n^2 \equiv 1 \text{ modulo } p,$$

implying $p = 2$.

If $N(\phi) = 4$, then ϕ is not of Type C'' . If ϕ were of Type C' , then we would infer from Theorem 1, (a) and Theorem 2C' (a) that

$$4 \equiv 1 + n + n^2 \equiv 3 \text{ modulo } p$$

which is impossible. Hence ϕ is of Type B; and it follows from Theorem 2B that

$$4 \equiv N(\phi) \equiv 1 + n + n^2 \equiv 1 \text{ modulo } p.$$

Hence $p = 3$, completing the proof of (4).

If $N(\phi) = 5$, then it follows from (1) that $5 \equiv 1, 3 \text{ modulo } p$, implying $p = 2$.

If $N(\phi) = 6$, and if ϕ is of Type B, then we deduce from Theorem 2B that $6 \equiv 1 + n + n^2 \equiv 1 \text{ modulo } p$ or $p = 5$; and if ϕ is of Type C, then ϕ is of Type C' , and it follows from Theorem 2C' that $6 \equiv 1 + n + n^2 \equiv 3 \text{ modulo } p$ or $p = 3$, completing the proof.

We conclude this section by giving a little information concerning the projectivities of Type D whose order is a power of a prime.

THEOREM 2D. *Suppose that ϕ is of prime power order $o(\phi) = p^m$, and that ϕ^{p^i} is of Type (D, j) .*

- (a) $p \neq 3$ implies $n \equiv j \text{ modulo } p^{i+1}$.
- (b) $N(\phi) \neq 0$ implies $n \equiv j \text{ modulo } p^{i+1}$.
- (c) If $p = 3$ and $N(\phi) = 0$, then ϕ^3 is of Types C'' or D and $n \equiv j \text{ modulo } 3^i$.

Proof. Since ϕ^{p^i} is of Type (D, j) , it possesses exactly $1 + j + j^2$ fixed points; and it follows from Theorem 1, (a) that

$$1 + j + j^2 \equiv N(\phi^{p^i}) \equiv 1 + n + n^2 \text{ modulo } p^{i+1}$$

or $(n - j)(n + j + 1) \equiv 0 \text{ modulo } p^{i+1}$. Consequently we have:

- (d) $n + j + 1 \not\equiv 0 \text{ modulo } p$ implies $n \equiv j \text{ modulo } p^{i+1}$.

Suppose, next, that $N(\phi) \neq 0$. Then there exists a fixed line w of ϕ . Clearly w is a fixed line of ϕ^{p^i} which carries exactly $j + 1$ fixed points of ϕ^{p^i} . Thus it follows from Theorem 1, (b) that $j + 1 \equiv N(\phi^{p^i}, w) \equiv n + 1 \text{ modulo } p^{i+1}$; and this implies (b).

Suppose, next, that $N(\phi) = 0$ and that $n + j + 1 \equiv 0 \text{ modulo } p$. Then we infer from Theorem 1, (a) that

$$1 + n + n^2 \equiv N(\phi) \equiv 0 \equiv 1 + n + j \text{ modulo } p$$

or $j \equiv n^2$ modulo p . If h is a fixed line of ϕ^{p^t} , then it follows from Theorem 1, (b) that

$$j + 1 \equiv N(\phi^{p^t}, h) \equiv n + 1 \text{ modulo } p$$

or

(e) $j \equiv n$ modulo p .

But we have shown before that $n^2 \equiv j$ modulo p . Hence $n^2 \equiv n$ modulo p or $n \equiv 0, 1$ modulo p . Since we verified that $0 \equiv 1 + n + n^2$ modulo p , $n \equiv 0$ modulo p is ruled out; and we find that

(f) $n \equiv 1$ modulo p .

Consequently $0 \equiv 1 + n + n^2 \equiv 3$ modulo p , proving $p = 3$. This completes the proof of (a).

Since $N(\phi) = 0$ and $N(\phi^{3^t}) \neq 0$, there exists a smallest integer k such that $N(\phi^{3^k}) \neq 0$. Clearly $0 < k$ and $N(\phi^{3^{k-1}}) = 0$.

Case 1. ϕ^{3^k} is of Type D.

Then there exists an integer j' , not less than 2, such that ϕ^{3^k} is of Type (D, j'); and it follows from (e), (f) that $1 \equiv n \equiv j'$ modulo 3. It follows from 2, Theorem 1, (a) that

$$0 \equiv N(\phi^{3^{k-1}}) \equiv N(\phi^{3^k}) \equiv 1 + j' + j'^2 \text{ modulo } 3^k.$$

But $j' = 1 + 3j''$ where $0 < j''$, since otherwise $j' < 2$. Hence

$$0 \equiv 1 + j' + j'^2 \equiv 3 + 9j''(1 + j'') \text{ modulo } 3^k.$$

From $3 \not\equiv 0$ modulo 9 we infer $k = 1$, so that in this case ϕ^3 is of Type D.

Case 2. ϕ^{3^k} is not of Type D.

If ϕ^{3^k} were of Types B or C', then we would infer from 1, Theorem 2 the existence of fixed elements of ϕ which is impossible. Since $N(\phi^{3^k}) \neq 0$, it follows now that ϕ^{3^k} is of Type C''; and it follows from Theorem 2C'', (c) that $N(\phi^3) = 3$, proving that in this case too $k = 1$ and that ϕ^3 is of Type C''.

Thus we have shown in both cases that $N(\phi^3) \neq 0$; and hence it follows from (b), applied upon ϕ^3 , that $n \equiv j$ modulo p^t , completing the proof.

4. Special groups of projectivities. If Δ is a group of projectivities,

then we denote by $o(\Delta)$ the order of the group Δ . Furthermore we say that an element x in the projective plane Π is Δ -invariant, if it is left invariant by every projectivity in Δ . (See 1, Remark 4.)

DEFINITION. *The group Δ of projectivities of the projective plane Π is a special group of projectivities, if the identity is the only projectivity of Type D in Δ .*

If Π is the projective plane over a Galois Field, then its groups of special collineations are special groups of projectivities in the meaning of the above Definition (see 3, Remark 5). The converse of this statement is, however, not true, as may be seen from simple examples.

THEOREM 1. *Suppose that Δ is a special group of projectivities. Then*

- (a) $o(\Delta)$ is a divisor of $(1 + n + n^2)(1 + n)n^3(n - 1)^2$.
- (b) *If there exist Δ -invariant points, then $o(\Delta)$ is a divisor of $(1 + n)n^3(n - 1)^2$.*
- (c) *If there exist different Δ -invariant points, then $o(\Delta)$ is a divisor of $n^2(n - 1)^2$.*
- (d) *If the Δ -invariant points are not collinear, then $o(\Delta)$ is a divisor of $(n - 1)^2$.*

*Proof.*¹⁶ The ordered quadruplet of points (R, S, T, U) will be termed an ordinary quadrangle, if no three of the four points R, S, T, U are collinear. The set H of ordinary quadrangles admits Δ , if $(R\phi, S\phi, T\phi, U\phi)$ is in H , whenever (R, S, T, U) is in H and ϕ is in Δ . We prove:

(i) *If the set H of ordinary quadrangles admits the special group Δ of projectivities, then $o(\Delta)$ is a divisor of the number of quadrangles in H .*

Suppose that ϕ is a projectivity in Δ and (M, N, P, Q) an ordinary quadrangle in H such that $(M, N, P, Q) = (M\phi, N\phi, P\phi, Q\phi)$. Then ϕ is of Type D, as has already been pointed out when introducing, in Section 1, the types. But Δ is special and thus it follows that $\phi = 1$. Consequently every quadrangle in H is mapped by the projectivities in Δ upon exactly $o(\Delta)$ distinct quadrangles in H and these sets are mutually exclusive, proving (i).

¹⁶ The author is much indebted to the referee for pointing out this elegant proof which is much simpler than the author's original one.

Suppose that the Δ -invariant points are not collinear. Then there exist three points P, Q, R which are not collinear and which are Δ -invariant. Consider the set $[P, Q, R]$ of all the ordinary quadrangles of the form (P, Q, R, X) . This set contains $(n-1)^2$ quadrangles, since there exist just $(n-1)^2$ points in Π which are neither on $P+Q$ nor on $Q+R$ nor on $R+P$ —there are just $3n$ points on these three lines. This set $[P, Q, R]$ admits Δ . Hence it follows from (i) that $o(\Delta)$ is a divisor of $(n-1)^2$, proving (d).

Assume, next, that there exist two different Δ -invariant points, say $P \neq Q$. Consider the set $[P, Q]$ of all the ordinary quadrangles of the form (P, Q, X, Y) . This set contains $n^2(n-1)^2$ quadrangles, since X may be selected as any one of the n^2 points, not on $P+Q$; and $[P, Q]$ admits Δ . As before we deduce from (i) that $o(\Delta)$ is a divisor of $n^2(n-1)^2$, proving (c).

Assume now that there exists at least one Δ -invariant point, say P . Consider the set $[P]$ of all the ordinary quadrangles of the form (P, X, Y, Z) . This set contains exactly $(n+n^2)n^2(n-1)^2$ quadrangles, since X may be selected as anyone of the $n+n^2$ points, not P . But $[P]$ admits Δ ; and thus (b) is a consequence of (i).

Consider, finally, the set of all the ordinary quadrangles (W, X, Y, Z) . Since W may be selected in $1+n+n^2$ different fashions, this set contains $(1+n+n^2)(1+n)n^2(n-1)^2$ different quadrangles. Since the set of all ordinary quadrangles admits Δ , our contention (a) is a consequence of (i), completing the proof.

THEOREM 2A. *Suppose that Δ is a special group of projectivities and that 3 is not a common divisor of $o(\Delta)$ and $1+n+n^2$. Then every projectivity, not 1, in Δ is fixed element free if, and only if, $o(\Delta)$ is a divisor of $1+n+n^2$.*

Proof. If the projectivities, not 1, in Δ are fixed element free, then every point P is mapped by the projectivities in Δ upon $o(\Delta)$ different points. Since these sets $P\Delta$ are mutually exclusive, it follows that $o(\Delta)$ is a divisor of the number of $1+n+n^2$ of points.

Suppose, conversely, that $o(\Delta)$ is a divisor of $1+n+n^2$ and that $\phi \neq 1$ is a projectivity in Δ . Then there exists an integer k such that ϕ^k is of order a prime number p . Since ϕ^k is in Δ , it is not of Type D and its order p is a divisor of $o(\Delta)$ and therefore of $1+n+n^2$. Hence $p \neq 3$; and it follows from 3, Theorem 3, (A) that ϕ^k is fixed element free. Consequently ϕ itself is fixed element free, completing the proof.

Whenever we speak of p -groups, this shall signify that p is a prime number and that the group under consideration is of order a power of p . The following proposition will be applied several times.

LEMMA 1. *If the group Δ of projectivities of Π is a p -group, if the set Ξ of elements in Π is mapped upon itself by the projectivities in Δ , then the number of Δ -invariant elements in Ξ is, modulo p , congruent to the number of elements in Ξ .*

Proof. If x is an element in Ξ , then consider the set $x\Delta$ of all the elements in Π upon which x is mapped by projectivities in Δ . Every set $x\Delta$, for x in Ξ , is part of Ξ ; and these sets are mutually exclusive. If x is Δ -invariant, then $x\Delta$ consists of one and only one element, namely x . If x is not Δ -invariant, then the number of elements in $x\Delta$ is a multiple of p —as a matter of fact a positive power of p . Our contention is an immediate consequence of these facts.

THEOREM 2B. *Suppose that Δ is a special group of projectivities.*

- (a) *Every projectivity, not 1, in Δ is of Type B if, and only if, $o(\Delta)$ is a divisor of n^3 .*
- (b) *If Δ is a p -group and p a divisor of n , then there exist Δ -invariant points (lines) and Δ -invariant points on every Δ -invariant line.*

Proof. Suppose first that every projectivity, not 1, in Δ is of Type B. If $o(\Delta)$ is divisible by the prime number p , then there exists, by Cauchy's Theorem, a projectivity ϕ of order p in Δ . Since ϕ is of Type B, we infer from 3, Theorem 3, (B), that p is a divisor of n . Thus every prime divisor of $o(\Delta)$ is prime to $(1 + n + n^2)(n^2 - 1)$; and it follows from Theorem 1, (a) that $o(\Delta)$ is a divisor of n^3 .

Assume, conversely, that $o(\Delta)$ is a divisor of n^3 . If $\phi \neq 1$ is a projectivity in Δ , then there exists a positive integer k such that ϕ^k is of order a prime number p . Since ϕ^k is in the special group Δ , it is not of Type D; and hence it follows from 3, Theorem 3, (B), that ϕ^k is of Type B as p is a divisor of $o(\Delta)$ and therefore of n . It follows now from 1, Corollary 1 that ϕ itself is of Type B, completing the proof of (a).

If p is a divisor of n , then $1 + n + n^2 \equiv 1$ modulo p . Thus it follows from Lemma 1, applied on the set of all the $1 + n + n^2$ points (lines), that the number of Δ -invariant points (lines) is congruent to 1 modulo p , proving

the first part of (b). If h is a Δ -invariant line, then it follows from Lemma 1, applied on the set of all the $n + 1$ points on h that the number of Δ -invariant points on h is congruent to $n + 1 \equiv 1$ modulo p , proving the existence of Δ -invariant points on h . This completes the proof.

Remark 1. It is clear from the proof of (a)—and may be deduced from Theorem 1, (a)—that $o(\Delta)$ is a divisor of n^3 , if each prime divisor of $o(\Delta)$ is a divisor of n .

COROLLARY 1. *If every projectivity, not 1, in Δ , is of Type B, and if there exist at least two Δ -invariant points (lines), then $o(\Delta)$ is a divisor of n^2 and all the projectivities, not 1, in Δ have the same axis (center).*

Proof. It is a consequence of Theorem 1, (b) and Theorem 2B, (a) that $o(\Delta)$ is a common divisor of n^3 and $n^2(n-1)^2$. Hence $o(\Delta)$ is a divisor of n^2 . The second contention is a consequence of the fact that the axis is the only fixed line of a projectivity of Type B which carries more than one fixed point.

In analogy to the division of cases used in 3, Theorem 3, we introduce now the following notation: the projectivity ϕ is said to be of *Type C**, whenever it is of Type C and $N(\phi) = 1$; and it is said to be of *Type C***, whenever it is of Type C though $N(\phi) \neq 1$.

THEOREM 2C*. *Suppose that Δ is a special group of projectivities and that $o(\Delta)$ and $n + 1$ are not both even.*

- (a) *Every projectivity, not 1, in Δ is of Type C* if, and only if, $o(\Delta)$ is a divisor of $n + 1$.*
- (b) *If Δ is nilpotent and $o(\Delta)$ a divisor of $n + 1$, then all the projectivities, not 1, in Δ have the same fixed point and the same fixed line.*

Proof. Suppose first that every projectivity, not 1, in Δ is of Type C*. Thus every projectivity $\phi \neq 1$ in Δ possesses one and only one fixed point $H(\phi)$, one and only one fixed line $h(\phi)$; and $H(\phi)$ is not on $h(\phi)$. Assume now that the prime number p is a divisor of $o(\Delta)$. There exists, by Cauchy's Theorem, a projectivity ϕ of order p in Δ . If p were 2, then it would follow from our hypothesis that n would be even; and it would follow from 3, Theorem 3, (B), that ϕ would be of Type B, an impossibility. Thus p is odd. But ϕ is of Type C*; and hence it follows from 3, Theorem 3, (C*),

that p is a divisor of $n + 1$. Consequently every prime divisor of $o(\Delta)$ is prime to $(1 + n + n^2)n(n - 1)$; and it follows from Theorem 1, (a) that $o(\Delta)$ is a divisor of $n + 1$.

Assume, conversely, that $o(\Delta)$ is a divisor of $n + 1$. If $\phi \neq 1$ is a projectivity in Δ , then there exists a positive integer k such that ϕ^k is of order a prime number p . Since ϕ^k is in the special group Δ , it is not of Type D; since p is, as a divisor of $o(\Delta)$, a divisor of $n + 1$, it follows from our hypothesis that $p \neq 2$ and from 3, Theorem 3, (C*), that ϕ^k is of Type C*. Hence it follows from 1, Theorem 2—and the fact that fixed elements of ϕ are also fixed elements of ϕ^k —that ϕ itself is of Type C*, completing the proof of (a).

If Δ is nilpotent, then Δ is the direct product of p -groups Δ_p . If $o(\Delta)$ is a divisor of $n + 1$, and if $\Delta_p \neq 1$, then the prime number p is a divisor of $n + 1$. This implies that $1 + n + n^2 \equiv 1$ modulo p ; and we infer from Lemma 1, applied on the set of all the $1 + n + n^2$ points, the existence of Δ_p -invariant points and lines. Since it follows from (a) that all the projectivities, not 1, in Δ are of Type C*, we infer the existence of one and only one common fixed point $P(p)$ (fixed line $h(p)$) of the projectivities, not 1, in Δ_p . If p and q are different prime numbers, and if Δ_p and Δ_q are both different from 1, then consider projectivities ϕ and γ of orders p and q respectively. Since $\phi\gamma = \gamma\phi$, it follows that both ϕ and γ are powers of $\phi\gamma$. But $\phi\gamma$ is in Δ and therefore of Type C*. It follows from 1, Theorem 2, that the only fixed point $P(p)$ of ϕ is the only fixed point of $\phi\gamma$; and the only fixed point $P(q)$ of γ is likewise the only fixed point of $\phi\gamma$. Hence $P(p) = P(q)$; and $h(p) = h(q)$ is seen in a similar fashion. From this fact our contention (b) is now easily deduced.

Remark 2. It is clear from the proof of (a)—and may be deduced from Theorem 1, (a)—that $o(\Delta)$ is a divisor of $n + 1$, if it is odd and each of its prime divisors is a divisor of $n + 1$.

THEOREM 2C.** *Suppose that Δ is a special p -group not 1, and that $p \neq 2, 3$. Then the following properties imply each other.*

- (a) *The Δ -invariant points are not collinear.*
- (b) *Every projectivity, not 1, in Δ is of Type C**.*
- (c) *p is a divisor of $n - 1$.*
- (d) *$o(\Delta)$ is a divisor of $(n - 1)^2$.*

Proof. If (a) is satisfied, and if $\phi \neq 1$, then there exist at least three non-collinear fixed points of ϕ : Since ϕ cannot be of Type D, as Δ is special, this implies that ϕ is of Type C**. Thus (a) implies (b). That (b) implies (c), may be inferred from 3, Theorem 3, (C); if (c) is satisfied, then p is prime to $(1+n+n^2)n^3(n+1)$, since $p \neq 2, 3$; and (d) is a consequence of Theorem 1, (a).

Assume finally the validity of (d). Then p is a divisor of $n-1$ and hence $1+n+n^2 \equiv 3$ modulo p . Since $p \neq 2, 3$, we may infer from Lemma 1 the existence of at least three different Δ -invariant points. If P and Q are two different Δ -invariant points, then there exist $n^2 \equiv 1$ modulo p points outside the line $P+Q$; and hence it follows from Lemma 1 that there exist Δ -invariant points, not on $P+Q$, proving the validity of (a) and completing the proof.

THEOREM 2C. *If Δ is a special group of projectivities whose order is prime to 3, then the following properties imply each other.*

- (a) $o(\Delta)$ is a divisor of $(n+1)(n-1)^2$.
- (b) Every prime divisor of $o(\Delta)$ is a divisor of n^2-1 .
- (c) Every projectivity, not 1, in Δ is of Type C.

Proof. It is obvious that (a) implies (b). If (b) is satisfied, then $o(\Delta)$ is prime to $(1+n+n^2)n^3$, since $o(\Delta)$ is prime to 3. But Δ is special. Hence it follows from Theorem 1, (a) that $o(\Delta)$ is a divisor of $(n+1)(n-1)^2$, showing that (a) and (b) are equivalent.

Assume now the validity of (c). If p is a prime divisor of $o(\Delta)$, then there exists, by Cauchy's Theorem, a projectivity ϕ of order p in Δ . Since $p \neq 3$, and since ϕ is, by hypothesis, of Type C' or C'', we infer from 3, Theorem 2C' and 3, Theorem 2C'', that p is a divisor of n^2-1 , showing that (b) is a consequence of (c).

Assume, finally, the validity of (b). If ϕ is a projectivity, not 1, in Δ , then there exists a positive integer k such that ϕ^k is of order a prime number p . Since p is a prime divisor of $o(\Delta)$, p is different from 3 and p is a divisor of n^2-1 . Since Δ is special and $\phi^k \neq 1$, it follows that ϕ^k is not of Type D. It follows from 3, Theorem 2A and 3, Theorem 2B, that ϕ^k is neither of Type A nor of Type B. Consequently ϕ^k is of Type C. But the order of ϕ is prime to 3, since the order of Δ is prime to 3. Hence it follows from 1, Corollary 4 that ϕ itself is of Type C, showing that (c) is a consequence of (b). This completes the proof.

COROLLARY 2. *If Δ is a special group of projectivities whose order is prime to 3, then the following properties imply each other.*

- (a) $o(\Delta)$ is a divisor of $(n+1)n^3(n-1)^2$.
- (b) Every prime divisor of $o(\Delta)$ is a divisor of $n(n^2-1)$.
- (c) Every projectivity in Δ possesses fixed elements.

Proof. It is clear that (a) implies (b). If (b) is valid, then $o(\Delta)$ is prime to $1+n+n^2$, since it is prime to 3. But Δ is special; and hence it follows from Theorem 1, (a) that $o(\Delta)$ is a divisor of $(n+1)n^3(n-1)^2$, showing the equivalence of (a) and (b).

Assume now the validity of (c). If p is a prime divisor of $o(\Delta)$, then there exists, by Cauchy's Theorem, a projectivity ϕ of order p in Δ . Since $p \neq 3$, and since ϕ is, by hypothesis, fixed element free, it follows from 3, Theorem 3, (A), that p is prime to $1+n+n^2$. Hence it follows from 3, Corollary 7, that p is a divisor of $(n-1)n(n+1)$, showing that (b) is a consequence of (c).

Assume, finally, the validity of (b). If ϕ is a projectivity, not 1, in Δ , then we distinguish two cases.

Case 1. n and $o(\phi)$ are relatively prime.

If p is a prime divisor of $o(\phi)$, then p is a divisor of $o(\Delta)$ so that $p \neq 3$ is a divisor of n^2-1 . But $o(\phi)$ is the order of the cyclic group generated by ϕ which group is special as a subgroup of Δ . Hence we may infer from Theorem 2C that ϕ is of Type C so that in particular ϕ possesses fixed elements.

Case 2. n and $o(\phi)$ are not relatively prime.

Then there exists a prime number q which is a common divisor of n and $o(\phi)$. There exists furthermore a positive integer i such that ϕ^i is of order q . Since ϕ^i , as an element in the special group Δ , is not of Type D, it follows from 3, Theorem 2B, that ϕ^i is of Type B. Then it follows from 1, Theorem 2, that center and axis of ϕ^i are fixed elements of ϕ .

Thus we have shown in all generality that (c) is a consequence of (b) completing the proof.

5. The group of special collineations. If Π is the projective plane over the Galois Field $\text{GF}(n)$, then we have pointed out in 3, Remark 5, that the

* group Z of special collineations of Π is a special group of projectivities whose order is $(1+n+n^2)(1+n)n^3(1-n)^2$. This shows the impossibility of improving the limit given in Theorem 1, (a).

If P is some point in Π and $Z(P)$ the group of all the projectivities in Z which leave invariant the point P , then the order of $Z(P)$ is $(1+n)n^3(1-n)^2$, showing the impossibility of improving the limit given in Theorem 1, (b) and Corollary 2, (a). But there does not exist any $Z(P)$ -invariant line, showing the impossibility of substituting in Corollary 2 for (c) the condition: "there exist Δ -invariant lines."

If P and Q are different points in Π and $Z(P, Q)$ the group of all the projectivities in Z which leave invariant P and Q , then the order of $Z(P, Q)$ is $n^2(n-1)$; but the only $Z(P, Q)$ -invariant elements are P, Q and $P+Q$, showing the impossibility of improving Theorem 1, (c).

If the points P, Q, R in Π are not collinear, and if $Z(P, Q, R)$ is the group of all the projectivities in Z which leave invariant P, Q and R , then this group is of order $(n-1)^2$, showing the impossibility of improving the limits given in Theorem 1, (d) and Theorem 2C**.

Singer (1) has shown that Z possesses a cyclic subgroup of order $1+n+n^2$ whose generator effects a cyclic permutation of the $1+n+n^2$ points in Π . Thus Theorem 2A gives a best limit. The method employed by Singer (1) may be used to construct a cyclic subgroup of order $n+1$ of Z , whose generator leaves invariant a point P and effects a cyclic permutation of the $n+1$ points on some line, not through P , showing the impossibility of improving the limit given in Theorem 2C*.

Consider finally the subgroup Γ of Z whose elements may be represented by matrices of the form

$$\begin{pmatrix} 1 & 0 & 0 \\ a & 1 & 0 \\ b & c & 1 \end{pmatrix}.$$

It is clear that Γ is of order n^3 and that the projectivities in Γ do not all have the same axis and/or center. Thus the limits given in Theorem 2B cannot be improved.

UNIVERSITY OF ILLINOIS,
URBANA, ILLINOIS.

BIBLIOGRAPHY

R. Baer

- (1) "Homogeneity of projective planes," *American Journal of Mathematics*, vol. 64 (1942), pp. 137-152.
- (2) "Projectivities with fixed points on every line of the plane," *Bulletin of the American Mathematical Society*, vol. 52 (1946), pp. 273-286.

R. D. Carmichael

- (1) *Introduction to the theory of groups of finite order*, Ginn & Co., 1937.

Marshall Hall

- (1) "Projective planes," *Transactions of the American Mathematical Society*, vol. 54 (1943), pp. 229-277.

N. Jacobson

- (1) *The theory of rings*, Mathematical Surveys, No. 2, 1943.

J. Singer

- (1) "A theorem in finite projective geometry and some applications to number theory," *Transactions of the American Mathematical Society*, vol. 13 (1938), pp. 377-385.

M. Steck

- (1) "Grundlegung einer Theorie der reellen Inzidenzabbildungen in endlichen projektiven Geometrien I, II," *Journal für die reine und angewandte Mathematik*, vol. 179 (1938), pp. 37-64, 148-192.

A. Speiser

- (1) *Theorie der Gruppen von endlicher Ordnung*, 3. Auflage, J. Springer, Berlin, 1937.

O. Veblen and J. W. Young

- (1) *Projective Geometry*, vol. 1, Ginn & Co., 1910.

THE SUM FORMULA OF EULER-MACLAURIN AND THE INVERSIONS OF FOURIER AND MÖBIUS.*

By AUREL WINTNER.

1. By further developing the considerations of [11], the present paper deals with the legitimacy of equidistant Riemannian evaluations of improper integrals and with its connections with both kinds of inversions mentioned in the title.

The connection with the second of these inversions, that of Möbius, can easily be hidden. Actually, it is the basis of the following theorem:

(α) If $f(x)$ is continuous on the half-open interval $0 < x \leq 1$, and if the limit

$$(1) \quad \lim_{\epsilon \rightarrow 0} \sum_{n\epsilon \leq 1} f(n\epsilon)$$

exists, then the improper integral

$$(2) \quad \int_{+0}^1 f(x) dx$$

is convergent (and has the same value as (1)).

This theorem, (α), sounds innocent enough. Actually, it contains the following theorem of Hardy and Littlewood [5]:

(α_0) Every series summable in Lambert's sense (L) is summable in Abel's sense (A).

That (α) is quite deep, follows from the implication (α) \rightarrow (α_0) and from the fact that, according to Hardy and Littlewood, (α_0) contains the prime number theorem (incidentally, the converse does not hold, since the proof depends on a refinement of the prime number theorem, that is, on the non-vanishing of $\zeta(s)$ on a certain open domain containing the line, $\sigma = 1$, of the prime number theorem).

In order to verify that (α) \rightarrow (α_0), the continuity of $f(x)$, assumed in (α), is sufficient. From quite another point of view, namely, from that of the heavy discontinuities compatible with the measurability of a function, it is of interest that (α) can be refined as follows:

* Received December 10, 1946.

(α^*) If $f(x)$ is L -integrable on every interval $\epsilon \leq x \leq 1$, where $\epsilon > 0$, and if the limit (1) exists, then the improper integral (2) is convergent and has the value (1).

The converse of (α) is obviously false (for a simple example, which can readily be refined, cf. [2], p. 230). That even the converse of the corollary, (α_0), of (α) is false, is shown by an arithmetical example of Hardy and Littlewood [6], pp. 265-269. What is true in the "converse" direction proves to be a criterion of the following type:

(β) If $f(x)$ is of bounded variation on every interval $\epsilon \leq x \leq 1$, where $\epsilon > 0$, and behaves, as $x \rightarrow 0$, so as to satisfy the restriction

$$(3) \quad \int_{\epsilon}^1 |df(x)| = o(\epsilon^{-1}),$$

then the convergence of the improper integral (2) implies that the limit (1) exists (and has the same value as (2)).

Corresponding to the implication $(\alpha) \rightarrow (\alpha_0)$, the last criterion, (β), implies the following (partial) converse of (α_0):

(β_0) Every absolutely (A)-summable series is (L)-summable.

Not even this corollary of (β) seems to occur in the literature. What is known is that particular case of (β_0) in which the series is absolutely (A)-summable for the trivial reason that the derivative of the Abelian generator of the series has a *monotone majorant* which is integrable (criterion of Ananda-Rao [1]; a simplified proof is given by Hardy and Littlewood [6], pp. 258-259). In fact, the situation is as follows:

A series $\sum a_n$ is called (A)-summable if $A(r) = \sum a_n r^n$ (is convergent when $r < 1$ and) tends to a limit, $A(1-0)$, as $r \rightarrow 1$. Clearly, this is equivalent to the convergence of the improper integral

$$\int_0^{1-0} A'(r) dr, \quad \text{where} \quad A' = dA/dr.$$

Correspondingly, J. M. Whittaker has called the series $\sum a_n$ absolutely (A)-summable if this integral is absolutely convergent, that is, if

$$\int_0^{1-0} |dA(r)| < \infty.$$

What the known particular case of (β_0) , referred to above, states is that $\sum a_n$ is (L) -summable if the condition required by the last formula line is replaced by the assumption

$$\int_0^{1-0} (\max_{0 \leq q \leq r} |A'(q)|) dr < \infty,$$

which, of course, is more strict than

$$(4) \quad \int_0^{1-0} |A'(r)| dr < \infty,$$

the absolute (A) -summability of $\sum a_n$.

Only the second of the inversions mentioned in the title, the inversion of Möbius, seems to be involved above. That the other inversion, that concerning Fourier transforms, can be involved just as well, is clear from the connections considered in [11]. In fact, the connection between the inversion formulae of Möbius and Fourier becomes manifest if the sum occurring in (1), a sum arrested at $x = 1$, is made an infinite sum. The latter is an Euler-Maclaurin expression and admits, therefore, of Poisson's Fourier analysis (in [11], only the arrested sums, leading to Fourier constants rather than to Fourier transforms, have been considered).

This Fourier analysis can be carried out very simply, by starting from a suitable formulation of the Euler-Maclaurin formula itself. In fact, the latter leads automatically to the function $x - [x]$, where $[x]$ denotes the greatest integer not exceeding x . All that will then be needed is the insertion of the Fourier series of the periodic function $x - [x]$ into the Euler-Maclaurin formula. What will be needed of the properties of this series,

$$(5) \quad \frac{1}{2} - \pi^{-1} \sum_{n=1}^{\infty} n^{-1} \sin 2\pi nx,$$

is that (5) is convergent (uniformly on every closed interval not containing an $x = [x]$); that

$$(6) \quad \text{the sum of (5) is } x - [x] \text{ if } x \neq [x], \text{ and } 0 \text{ if } x = [x];$$

finally that

$$(7) \quad \text{the partial sums of (5) are uniformly bounded.}$$

2. The whole theory centers about the identity claimed by the following remark:

(i) If $f(x)$ is of bounded variation on the half-line $x \geq 1$,

$$(8) \quad \int_1^{\infty} |df(x)| < \infty,$$

and if the value of $\lim_{x \rightarrow \infty} f(x)$ (which then exists) is

$$(9) \quad f(\infty) = 0,$$

then the limit

$$(10) \quad \lim_{x \rightarrow \infty} \left(\sum_{n \leq x} f(n) - \int_1^x f(u) du \right),$$

where $n = 1, 2, \dots$, exists and is equal to the value of

$$(11) \quad \int_1^{\infty} (x - [x]) df(x).$$

If $f(x)$ is discontinuous at an integer, $x = n$, then both factors occurring in the Stieltjes integral (11) are discontinuous at $x = n$, and so (11) must then be meant in the Lebesgue-Stieltjes sense, rather than as an improper Riemann-Stieltjes integral.

The actual content of (i) is just the Euler-Maclaurin sum formula, except that the above formulation avoids the usual conditions of smoothness. In fact, the usual wording presupposes not only the continuity of $f(x)$ but its absolute continuity as well; cf. [4]. The latter restriction would preclude the applicability of the Euler-Maclaurin rule to "explicit" cases of considerable interest. Actually, the proof becomes particularly simply precisely under the above general assumptions.

In fact, since $[x] - x = O(1)$ as $x \rightarrow \infty$, and since (9) means that $f(x) = o(1)$, a partial integration gives

$$(12) \quad \int_1^x f(u) d([u] - u) = o(1)O(1) - 0 - \int_1^x ([u] - u) df(u).$$

But $[u] - u = O(1)$ and (8) imply that, as $x \rightarrow \infty$, the expression on the right of (12) tends to the value (11), whereas the integral on the left of (12) is identical with the difference the limit of which is taken in (10). This proves (i).

(ii) If $f(x)$ is defined for $x > 1$ in such a way that the Lebesgue-Stieltjes integral of $u - [u]$ with respect to $f(u)$ exists on every finite

interval $1 \leq u \leq x$ and tends, as $x \rightarrow \infty$, to a finite limit, (11), and if $f(x) \rightarrow 0$, then the limit (10) exists and equals the value of (11).

This generalization of (i) is clear from the proof of (i). In fact, (12) can be applied under the present assumptions also.

An immediate consequence of (i) is the following fact:

(iii) If $f(x)$ is of bounded variation on every closed half-line contained in the open half-line $x > 0$, and if the improper integral of $f(x)$ on such a closed half-line is convergent (possibly just conditionally), then the series $\sum f(\epsilon n)$ is convergent and satisfies the inequality

$$(13) \quad \left| \epsilon \sum_{n=1}^{\infty} f(\epsilon n) - \int_{\epsilon}^{\infty} f(x) dx \right| \leq \epsilon \int_{\epsilon}^{\infty} |df(x)|,$$

where $\epsilon > 0$ is arbitrary.

Easy examples show that the convergence of both

$$(14) \quad \int_1^{\infty} |df(x)| \quad \text{and} \quad \int_1^{\infty} f(x) dx$$

does not preclude the divergence of

$$(15) \quad \int_1^{\infty} |f(x)| dx;$$

so that the parenthetical remark preceding (13) is not illusory.

Since the convergence of the first of the integrals (14) implies the existence of a limit $f(\infty)$, and since the non-vanishing of the latter is prevented by the convergence of the second of the integrals (14), the assumptions of (iii) contain those of (i). In addition, (i) is now applicable to $f(\epsilon x)$ instead of to $f(x)$, if $\epsilon > 0$ is arbitrarily fixed. Thus, (10) being equal to (11),

$$(16) \quad \lim_{x \rightarrow \infty} \left(\sum_{n \leq x/\epsilon} f(\epsilon n) - \int_1^{x/\epsilon} f(\epsilon u) du \right) = \int_1^{\infty} (x - [x]) d_x f(\epsilon x).$$

If this is multiplied by ϵ , then, since the second of the integrals (14) is convergent, what results on the left of (16) is the difference the absolute value of which is taken on the left of (13). Since the integral on the right of (16) is majorized by

$$\int_1^{\infty} (x - [x]) |df(x)| \leq \int_1^{\infty} |df(x)| = \int_{\epsilon}^{\infty} |df(x)|,$$

the assertion, (13), of (iii) follows.

(iv) *If both integrals (14) are convergent and*

$$(17) \quad \int_{\epsilon}^1 |df(x)| = o(\epsilon^{-1})$$

as $\epsilon \rightarrow 0$, then

$$(18) \quad \epsilon \sum_{n=1}^{\infty} f(n\epsilon) - \int_{\epsilon}^{\infty} f(x) dx \rightarrow 0.$$

This is a corollary of (iii). It should be noted that neither of the terms on the left of (18) need tend to a limit. All that follows is that, if one of these terms tends to a limit, the other must tend to the same limit. Thus (iv) implies that *the Riemannian equidistant evaluation* of a convergent (doubly improper) integral

$$(19) \quad \int_0^{\infty} f(x) dx = \int_0^1 + \int_1^{\infty}, \quad \text{where } \int_0^1 = \lim_{\epsilon \rightarrow 0} \int_{\epsilon}^1, \quad \int_1^{\infty} = \lim_{x \rightarrow \infty} \int_1^x,$$

is legitimate whenever

$$(20) \quad \int_{\epsilon}^{\infty} |df(x)| = o(\epsilon^{-1}).$$

(v) *If $f(x)$ has on every half-line $x > \epsilon$ a finite total variation which satisfies (20) as $\epsilon \rightarrow 0$, and if both improper integrals occurring in (19) are convergent (possibly just conditionally), then*

$$(21) \quad \epsilon \sum_{n=1}^{\infty} f(n\epsilon) \rightarrow \int_0^{\infty} f(x) dx.$$

If $f(x)$ is chosen to be 0 when $x > 1$, and if the continuous variable ϵ is replaced by the reciprocal value of an integer, it follows that, for every function $f(x)$ which is defined on the interval $0 < x < 1$ so as to satisfy (17), the (equidistant) Riemannian relation

$$(22) \quad \sum_{k=1}^n f(k/n)/n \rightarrow \lim_{\epsilon \rightarrow 0} \int_{\epsilon}^1 f(x) dx \quad (n \rightarrow \infty)$$

holds whenever the improper integral on the right of (21) is convergent. This corollary of (v) is sharper than the usual criterion for the truth of (22); a criterion which replaces (17) by the monotony of $f(x)$ (cf. [2], pp. 229-230). For, on the one hand, (17) and the convergence of the improper integral on the right of (21) do not imply the absolute convergence of that integral. And, on the other hand, the convergence of the latter and the monotony of $f(x)$ imply the absolute convergence of the integral, which in turn implies, again by the monotony of $f(x)$, that

$$(23) \quad \epsilon f(\epsilon) \rightarrow 0 \text{ as } \epsilon \rightarrow 0,$$

and so (17) follows by necessity.

The criteria (iii), (iv), (v) do not take care of cases exemplified by the function $f(x) = (\sin x)/x$. Functions of this type are included in the following theorem:

(vi) *If λ is a non-vanishing real number, then the limit relation*

$$(24) \quad \epsilon \sum_{n=1}^{\infty} f(n\epsilon) e^{i\lambda n\epsilon} \rightarrow \int_0^{\infty} f(x) e^{i\lambda x} dx \quad (\epsilon \rightarrow 0)$$

holds whenever

$$(25) \quad \int_0^{\infty} |df(x)| < \infty \quad \text{and } f(\infty) = 0$$

(the convergence of the series and of the integral occurring in (24) being implied by (25), since $\lambda \neq 0$).

The truth of the parenthetical remark following (25) is seen after a partial summation and a partial integration of the series (24) and of the integral (24), respectively.

The proof of (vi) may be omitted, since it can be read off from one given by Bromwich and Hardy [2], pp. 231-233. It is true that they assume, instead of (25), that $f(x)$ is a *monotone* function satisfying (25), and it is also true that not every real-valued $f(x)$ satisfying (25) is the sum of two monotone functions both of which satisfy (25). Nevertheless, a glance at

the proof given by Bromwich and Hardy shows that (25) alone suffices for the proof of (24).

Incidentally, the proof of (vi) differs from that of the preceding criteria only insofar as what corresponds to an application of the second mean-value theorem (namely, preliminary partial summations and partial integrations of the respective sums and integrals) must precede an application of the first mean-value theorem.

3. If

$$(26) \quad \mu^*(x) = \sum_{n=1}^x \mu(n)/n, \quad \left(\sum_{n=1}^x = \sum_{n \leq x} \right),$$

where $\mu(n)$ denotes Möbius' factor, then the prime number theorem is known to be equivalent to $\mu^*(x) = o(1)$, where $x \rightarrow \infty$. In their proof of the theorem denoted above by (α_0) , Hardy and Littlewood [5] refer to somewhat more than $\mu^*(x) = o(1)$, namely, to $\mu^*(x) = O(\log x)^{-2}$. What is actually needed is something between these two estimates, namely,

$$(27) \quad \sum_{n=1}^{\infty} |\mu^*(n)|/n < \infty.$$

In order to prove that the theorem denoted above by (α^*) can be concluded from (27), it is sufficient to repeat the proof of (α_0) . The same proof also supplies the following extension of (α^*) :

(vii) *Let $f(x)$ be L -integrable on every closed half-line contained in the open half-line $x > 0$, and suppose that $f(x)$ tends quite rapidly to 0, as $x \rightarrow \infty$; for instance, so that*

$$(28) \quad f(x) = O(x^{-1-\eta}) \text{ as } x \rightarrow \infty$$

holds for some $\eta > 0$. Then the limit

$$(29) \quad \lim_{\epsilon \rightarrow 0} \epsilon \sum_{n=1}^{\infty} f(n\epsilon)$$

cannot exist unless the improper integral

$$(30) \quad \int_{+0}^{\infty} f(x) dx = \lim_{\epsilon \rightarrow 0} \int_{\epsilon}^{\infty} f(x) dx$$

is convergent, and has the same value as (29).

Needless to say, the convergence of the series occurring in (29) and of the integral occurring on the right of (30) is assured by (28) for every $\epsilon > 0$. If the convergence, or the absolute convergence, of these expressions (for every $\epsilon > 0$) is granted, one has the impression that the actual rôle of (28) is that of a "Tauberian condition." But this is not quite the case. In fact, something like (28) will be needed twice, both times at very *rough* stages of the proof of (vii); namely, in order to make Möbius' inversion legitimate at a fixed $\epsilon > 0$, that is, *before* the application of the crucial limit process, $\epsilon \rightarrow 0$, and then, in order to justify a term-by-term integration, but again just for a fixed $\epsilon > 0$. Nevertheless, (28) cannot be replaced by

$$(31) \quad \sum_{n=1}^{\infty} |f(n\epsilon)| < \infty \text{ and } \int_{\epsilon}^{\infty} |f(x)| dx < \infty,$$

where $\epsilon > 0$ is arbitrary. For, if (28) could be relaxed to (31) in (vii), it would be clear from the proof of (vii) (cf. below) that not only (α_0) but also the converse of (α_0) is true. However, the converse of (α_0) is disproved by the example of Hardy and Littlewood, mentioned after (α^*) . The situation is cleared up by the fact that, as proved in [11], pp. 17-18, the absolute convergence of the series (31) is insufficient for the legitimacy of Möbius' inversion (at a fixed $\epsilon > 0$).

In the proof of (vii), it can be assumed that the limit (29), the existence of which is assumed, is 0. For, if it is not 0, it can be made 0 by an alteration of $f(x)$ on an unessential range, say on the interval $1 < x < 2$. Thus the assumptions of (vii) are reduced to (28) and

$$(32) \quad F(x) = o(x^{-1}) \text{ as } x \rightarrow 0,$$

where $x > 0$ and

$$(33) \quad F(x) = \sum_{n=1}^{\infty} f(nx),$$

whereas the assertion of (vii) becomes

$$(34) \quad \lim_{x \rightarrow 0} \int_x^{\infty} f(u) du = 0;$$

cf. (29) and (30).

Möbius' inversion of (33) is

$$(35) \quad f(x) = \sum_{n=1}^{\infty} \mu(n) F(nx),$$

where $x > 0$ is arbitrary. The legitimacy of this inversion is guaranteed by (28); in fact, somewhat less than (28) is sufficient (cf. [11], pp. 16-17). It is also seen from (33) and (28) that term-by-term integration of (35) on every closed half-line $x \geq \epsilon$, where $\epsilon > 0$, is legitimate:

$$(36) \quad \int_x^\infty f(u) du = \sum_{n=1}^\infty \mu(n) \int_x^\infty F(nu) du.$$

Hence, the assertion, (34), is equivalent to

$$(37) \quad \lim_{x \rightarrow 0} \sum_{n=1}^\infty \mu(n) \int_x^\infty F(nu) du = 0.$$

But (37) follows from (26), (27) and (32) by the same partial summation which Hardy and Littlewood [5] apply in their proof of the implication $(L) \rightarrow (A)$, that is, of (α_0) .

It should be emphasized that, due to the first part of (27), this proof is purely "Abelian" in nature. Correspondingly, (vii) is beyond the scope of the "Tauberian" technique of Karamata-Weiner, which would involve just the prime number theorem, that is, the second part of (27). In fact, this methodical situation prevails not only for (vii) but for (α_0) as well, and (α_0) is contained in (vii).

4. Before combining the facts preceding (vii) with the Fourier expansion, (5), of the kernel, $x - [x]$, of the Euler-Maclaurin integral, (11), it is convenient to replace (i) by the following variant of (ii):

(viii) *Let $f(x)$ be defined on the open half-line $x > 0$ in such a way that*

$$(38) \quad f(x) \rightarrow 0 \text{ as } x \rightarrow \infty$$

and

$$(39) \quad \epsilon f(\epsilon) \rightarrow 0 \text{ as } \epsilon \rightarrow 0,$$

and that the Lebesgue-Stieltjes integral of $u - [u]$ with respect to $f(u)$ exists on every closed, bounded interval, $\epsilon \leq u \leq x$. Suppose further that both improper integrals

$$(40) \quad \int_{+\infty}^{1-0} u df(u) = \lim_{\epsilon \rightarrow 0} \int_\epsilon^{1-0}, \quad \int_1^\infty (u - [u]) df(u) = \lim_{x \rightarrow \infty} \int_1^x$$

are convergent (possibly just conditionally). Then the limit

$$(41) \quad \lim_{x \rightarrow \infty} \left(\sum_{n \leq x} f(n) - \int_{+0}^x f(u) du \right),$$

where $n = 1, 2, \dots$, exists and is equal to the value of

$$(42) \quad \int_{+0}^{\infty} (u - [u]) df(u),$$

the sum of the two integrals (40).

In order to see this, it is sufficient to observe that $u - [u] = u$ when $0 < u < 1$, and that what therefore corresponds to (12) is

$$\int_{\epsilon}^x f(u) d([u] - u) = o(1)O(1) - o(1) - \int_{\epsilon}^x ([u] - u) df(u)$$

as $(\epsilon, x) \rightarrow (+0, \infty)$; cf. (38) and (39).

(ix) Let $f(x)$ be of finite total variation on every closed half-line contained in the open half-line $x > 0$. For $x \rightarrow \infty$, choose $f(\infty) = 0$. For $x \rightarrow +0$, suppose that

$$(43) \quad \int_{\epsilon}^{\infty} |df(x)| = o(\epsilon^{-1}) \text{ as } \epsilon \rightarrow 0$$

and that the improper integral

$$(44) \quad \int_{+0}^1 f(x) dx \text{ is convergent}$$

(possibly just conditionally). Then, if $t > 0$, the limit

$$(45) \quad F^*(t) = \lim_{x \rightarrow \infty} \left(t \sum_{n \leq x} f(nt) - \int_{+0}^x f(u) du \right),$$

where $n = 1, 2, \dots$, exists and has the value of

$$(46) \quad F^*(t) = t \int_{+0}^{\infty} (x/t - [x/t]) df(x),$$

and the improper integral

$$(47) \quad f^*(t) = \int_{+0}^{\infty} f(x) \cos tx \, dx = \lim_{\epsilon \rightarrow 0} \int_{\epsilon}^1 + \lim_{x \rightarrow \infty} \int_1^x$$

is convergent and represents a continuous function of $t > 0$.

First, the assumptions of (ix) imply those of (viii). This is clear, except for the assumption (39) of (viii). But (39) follows from the remarks made in connection with (23). On the other hand, if $t > 0$ is fixed, the assumptions of (ix) are satisfied for $f(tx)$ if they are satisfied for $f(x)$. Hence, the assumptions of (ix) imply that (viii) is applicable to $f(tx)$. This proves the existence of the limit (45), and the representation (46) of this limit.

Next, since $\sin(\epsilon t)$ is $O(\epsilon)$ when t is fixed, (38), (44) and (8) imply that the improper integral

$$(48) \quad \int_{+0}^{\infty} \sin tx \, df(x)$$

is convergent. On the other hand, if the integration range, $(+0, \infty)$, of (48) is replaced by the interval (ϵ, x) , then, if $t \neq 0$, a partial integration, when followed by an application of (38) and (39), shows that the resulting approximation to (48) tends, as $(\epsilon, x) \rightarrow (+0, \infty)$, to a limit, and that the latter is $-t$ times the integral (47). Hence, in order to complete the proof of (ix), only the continuity of the function (47), where $t \neq 0$, remains to be ascertained. But this is clear from the fact that the preceding limit process is uniform on every closed, bounded t -interval not containing $t = 0$.

(x) If $f(x)$ is a function of finite total variation on the half-line $x > 0$ and tends to 0 as $x \rightarrow \infty$ (that is, if the assumptions (43), (44) of (ix) are replaced by the stricter assumption

$$(49) \quad \int_{+0}^{\infty} |df(x)| < \infty$$

and $f(\infty) = 0$ is retained), then the series

$$(50) \quad \sum_{n=1}^{\infty} f^*(nt), \text{ where } f^*(t) = \int_0^{\infty} f(x) \cos tx \, dx,$$

is convergent for every $t > 0$ and is connected with the Euler-Maclaurin function (45) by the relation

$$(51) \quad F^*(t) = -\frac{1}{2}tf(+0) + 2 \sum_{n=1}^{\infty} f^*(2\pi n/t), \quad (t > 0),$$

provided that $f(x)$ is normalized by

$$(52) \quad 2f(x) = f(x+0) + f(x-0), \quad (x > 0).$$

This normalization, made possible by (49), has no influence on the integrals occurring in (44) and (50).

The identity (51) is just Poisson's Fourier analysis (cf. [7], pp. 78-79) of the Euler-Maclaurin sum formula, first proved, and found independently, by Dirichlet. Under the above assumptions, it is an immediate consequence of (6) and (7).

First, if the value of (5) were always $\phi(x) = x - [x]$, then $\phi(x/t)$ could be substituted from (5) into (46). According to (6), the value of (5) ceases to be $\phi(x)$ when x is an integer. But this discrepancy has an effect on the Stieltjes integral (46) only at points at which f , too, is discontinuous. Hence, the discrepancy has no effect if f has no discontinuities at all. On the other hand, when f has jumps, then, since the latter are normalized by (52), the discrepancy is compensated by the fact that, if $\phi_0(x)$ denotes the value of (5), the normalization (52) holds, by (6), for $f = \phi_0$ also. Accordingly, the function, $\phi(x/t)$, which multiplies $df(x)$ in (46), can be replaced by $\phi_0(x/t)$.

Since (7) and (49) make it clear that term-by-term integration of the series $\phi_0(x/t)$ is legitimate in (46), it now follows from the expansion (5) that

$$F^*(t) = t \int_{+0}^{\infty} \frac{1}{2} df(x) - t\pi^{-1} \sum_{n=1}^{\infty} n^{-1} \int_{+0}^{\infty} \sin(2\pi nx/t) df(x).$$

But the first integral on the right is $-\frac{1}{2}f(+0)$, by (38); and, what concerns the second integral, the function (48) was seen to be identical with $-t$ times the function (47). This gives

$$F^*(t) = -\frac{1}{2}tf(+0) - t\pi^{-1} \sum_{n=1}^{\infty} n^{-1} f^*(2\pi n/t) (-2\pi n/t),$$

which is (51).

(xi) If $f(x)$ is of finite total variation on the half-line $x > 0$ and if the improper integral

$$(53) \quad f^*(0) = \int_0^{\infty} f(x) dx = \lim_{x \rightarrow \infty} \int_0^x$$

is convergent, then

$$(54) \quad \sum_{n=1}^{\infty} f^*(nt) \rightarrow 0 \text{ as } t \rightarrow \infty,$$

where $f^*(t) = \int_0^{\infty} f(x) \cos tx dx$, whereas

$$(55) \quad t \sum_{n=1}^{\infty} f^*(nt) \rightarrow \frac{1}{2}\pi f(+0) \text{ as } t \rightarrow +0;$$

and, if $t > 0$,

$$(56) \quad t^{\frac{1}{2}} \left\{ \frac{1}{2}f(+0) + \sum_{n=1}^{\infty} f(nt) \right\} = t^{-\frac{1}{2}} \{ f^*(0) + 2 \sum_{n=1}^{\infty} f^*(2\pi n/t) \}.$$

The last identity is Cauchy's celebrated formula of reciprocity. It follows from (51) and (45), since the assumptions of (xi) imply those of (x), the convergence of (53) being assumed now. In fact, this assumption and (49) imply (38).

The proof of (54) can be based on the remark that (56) transforms (54) into

$$\epsilon \sum_{n=1}^{\infty} f(n\epsilon) \rightarrow f^*(0), \quad \epsilon \rightarrow 0$$

(as seen if (56) is multiplied by $t^{\frac{1}{2}}$ and t is then identified by ϵ , finally the $2\pi/t$, occurring on the right of (55), is replaced by ϵ). In view of the definition, (53), of $f^*(0)$, the assertion of the last formula line is equivalent to (18). But (18) is applicable, since the assumptions of (xi) imply those of (iv). This proves (54).

The remaining assertion, (55), is a relation of Gibbs' type (in this regard, cf. [11], p. 5 and p. 28). It can be obtained by observing that, according to (49) and (53),

$$\sum_{n=1}^{\infty} f(nt) \rightarrow 0, \quad t \rightarrow \infty.$$

In view of (56), this implies that

$$\frac{1}{2}f(+0) + o(1) = O(t^{-1}) + 2t^{-1} \sum_{n=1}^{\infty} f^*(2\pi n/t)$$

or, if t is replaced by $2\pi t$,

$$\frac{1}{2}f(+0) + o(1) = \pi^{-1}t^{-1} \sum_{n=1}^{\infty} f^*(n/t),$$

as $t \rightarrow \infty$. Since the last relation is equivalent to (55), the proof of (xi) is complete.

The limit relation (54) neither contains, nor is contained in, the estimate

$$(57) \quad |f^*(t)| < \text{const.}/t \quad (t > 0).$$

The latter holds even if the assumptions of (xi) are relaxed to those of (x). For, on the one hand, (48) is majorized by (49) and, on the other hand, (48) is $-t$ times the function (47).

The other limit relation, (55), can be interpreted as follows:

(xii). If $f(x)$ satisfies the assumptions of (xi), then its Fourier cosine transform, $f^*(t)$, satisfies the equidistant Riemannian relation,

$$(58) \quad \epsilon \sum_{n=1}^{\infty} f^*(n\epsilon) \rightarrow \int_{+0}^{\infty} f^*(t) dt \text{ as } \epsilon \rightarrow 0$$

(the convergence of the improper integral

$$(59) \quad \int_{+0}^{\infty} f^*(t) dt = \lim_{\epsilon \rightarrow 0} \int_{\epsilon}^1 + \lim_{x \rightarrow \infty} \int_1^x,$$

as well as the convergence of the series on the left of (58), is part of the statement).

In order to see this, let the function $f(x)$, given for $x > 0$, be extended for all x by placing $f(x) = 0$ when $x < 0$, and $f(0) = \frac{1}{2}f(+0)$. Then the assumptions of (xi) imply that

$$(60) \quad \int_{-\infty}^{\infty} |df(x)| < \infty$$

and

$$(61) \quad f(x) \rightarrow 0 \text{ as } x \rightarrow \pm \infty,$$

and that (52) holds for $-\infty < x < \infty$. But Pringsheim ([10], pp. 405-406; cf. also Hahn's presentation in [3], pp. 318-321) has shown that, if $f(x)$ satisfies (60), (61) and (52), then its representation by Fourier's "double integral" is valid for every x , provided that the exterior integration is interpreted as improper not only at ∞ but at 0 as well. If this is applied at $x = 0$ to the above extension of the given function $f(x)$, $0 < x < \infty$, then, since $f^*(t)$ is defined by (47), it follows that both parts of the improper integral (59) converge, and that the sum (59) has the value $\pi f(0)$.

The assumption that the integral (53) converges has not been used thus far. If it is used, (xi) becomes applicable, and so (58) follows from (55), where $\frac{1}{2}\pi f(+0) = \pi f(0)$.

Crucial in this deduction is the convergence of the improper integral (59) (along with the circumstance that (59) actually is improper at $t = 0$, in general), as is, correspondingly, that assumption of (xi) which is not required in (x), namely, the convergence of the improper integral (53). It is instructive to compare this situation with the proof of the following fact, which is more on the surface:

(xiii) If $f(x)$ satisfies just the assumptions of (x) (so that the additional restriction of (xi) and (xii), that requiring the convergence of (53), is not assumed), then, as $\epsilon \rightarrow 0$,

$$(62) \quad \epsilon \sum_{n=1}^{\infty} f(n\epsilon) \cos (tn\epsilon) \rightarrow f^*(t), \text{ where } f^*(t) = \int_0^{\infty} f(x) \cos tx \, dx,$$

if $t \neq 0$ (and

$$(62 \text{ bis}) \quad \epsilon \sum_{n=1}^{\infty} f(n\epsilon) \sin t(n\epsilon) \rightarrow \int_0^{\infty} f(x) \sin tx \, dx,$$

where $t = 0$ need not, of course, be excluded).

In fact, (xiii) is a mere restatement of (vi), where $0 \neq \lambda = \pm t$. The case $0 = \lambda = t$ of (24), (62), excluded in (vi), (xiii), represents the more delicate assertion of (xii).

5. A sufficient criterion for the legitimacy of Möbius' inversion (and for the inversion of this inversion, which is not always the same thing; cf. [11], pp. 17-18) is as follows: If

$$(63) \quad x_n = O(n^{-1-\eta})$$

is required for some $\eta > 0$, then the infinity of equations

$$(64) \quad \sum_{m=1}^{\infty} x_{nm} = y_n$$

for the infinity of unknowns x_1, x_2, \dots has the unique solution

$$(65) \quad x_n = \sum_{m=1}^{\infty} \mu(m) y_{nm},$$

(no matter what the sequence of data y_1, y_2, \dots is), where μ denotes Möbius' factor; whereas, if (63) is omitted and

$$(66) \quad y_n = O(n^{-1-\delta})$$

is required for some $\delta > 0$, then (65), when considered as a system for the unknowns y_1, y_2, \dots has a unique solution, which is precisely the y_n supplied by (64) (no matter what the sequence of data x_1, x_2, \dots , given in (65), is this time).

Somewhat less (but not very much less; cf. [11], pp. 16-18) than the assumption of an $\eta > 0$ or a $\delta > 0$ is sufficient. The proof requires, of course, nothing but formal rearrangements which, under the respective restrictions

(63), (64), are justified by the absolute convergence of the repeated summations involved.

It should be noted that, by virtue of (64) and (65), *the existence of an $\eta > 0$ is equivalent to the existence of a $\delta > 0$* , and that, as a matter of fact,

$$(67) \quad \text{l. u. b. } \eta = \text{l. u. b. } \delta,$$

if either of the Dedekind cuts occurring in (67) is positive (possibly ∞). This is seen by substituting (63) into (64), and (66) into (65), where $\mu(m) = O(1)$.

The symmetry relation (67) will be used below to an end corresponding to the one-to-one correspondences, developed in [11], pp. 25-26, between the smoothness of a periodic R -integrable function and the rapidity of the convergence of its equidistant R -sums.

For the sake of convenient references, let the preceding remarks be summarized in a slightly different form, as follows:

(*) *If t varies on the half-line $t > 0$, and if either*

$$(68) \quad g(t) = O(t^{-1-\eta}), \quad t \rightarrow \infty,$$

holds for some $\eta > 0$ or

$$(69) \quad h(t) = O(t^{-1-\delta}), \quad t \rightarrow \infty,$$

holds for some $\delta > 0$, then

$$(70) \quad h(t) = \sum_{n=1}^{\infty} g(nt), \quad 0 < t < \infty,$$

is equivalent to

$$(71) \quad g(t) = \sum_{n=1}^{\infty} \mu(n)h(nt), \quad 0 < t < \infty,$$

and, by virtue of the reciprocal relations (69), (70), either of the restrictions (68), (69) implies the other restriction; in fact, even the "best" values of the exponents η , δ are identical (in the sense of (67), where l. u. b. > 0 is assumed and l. u. b. $= \infty$ is allowed).

In order to verify this criterion, (*), it is sufficient to replace t in (70) and (71) by kt , where $k = 1, 2, \dots$; to identify the resulting systems with (60), (65) (for a fixed t), and (68), (69) with (63), (66), respectively;

finally, to particularize the parametric integer ($=k$) to be 1 in the resulting inversions.

The *self-reciprocal* relationship expressed by (69), (68) and (67) has variants on other scales; for instance, on the following exponential scale:

(* bis) *The assertions of (*) remain true if (67) is retained but (68) and (69) are replaced by*

$$(70 \text{ bis}) \quad g(t) = O(\exp - t^\eta), \quad t \rightarrow \infty$$

and

$$(71 \text{ bis}) \quad h(t) = O(\exp - t^\delta), \quad t \rightarrow \infty,$$

respectively, where $\eta > 0$ and $\delta > 0$.

In fact, since the assumptions of (* bis) imply those of (*), it is sufficient to ascertain that (67) holds in the case of (* bis) also. But this follows in the same way as before; namely, by substituting the O -assumptions into the respective series (70), (71), where $\mu(n) = O(1)$, as $n \rightarrow \infty$, and then using the fact that, as $t \rightarrow \infty$, both estimates

$$(72) \quad \sum_{n=1}^{\infty} (nt)^{-1-\epsilon} = O(t^{-1-\epsilon}), \quad (73) \quad \sum_{n=1}^{\infty} \exp - (nt)^\epsilon = O(\exp - t^\epsilon),$$

the first of which belongs to (*) and the second to (* bis), hold for every fixed $\epsilon > 0$.

These Möbius criteria will now be combined with the consequences of the Euler-Maclaurin lemma (ix) and its Fourier corollaries, deduced above from (5), (6), (7). The simplest fact which thus results can be isolated as follows:

(xiv) *If $f(x)$, where $x > 0$, satisfies the conditions*

$$(74) \quad \int_0^{\infty} |df(x)| < \infty \text{ and } f(\infty) = 0,$$

and if $f^*(t)$, $F^*(t)$ denote the functions which (47), (45) then define for $t > 0$, the existence of an α satisfying

$$(75) \quad \alpha > 1 \text{ and } f^*(t) = O(t^{-\alpha}) \text{ as } t \rightarrow \infty$$

is necessary and sufficient for the existence of a β satisfying

$$(76) \quad \beta > 1 \text{ and } F^*(x) = -\frac{1}{2}f(+0)x + O(x^\beta) \text{ as } x \rightarrow +0;$$

in fact, (74) and either of the assumptions (75), (76) imply that

$$(77) \quad \alpha^* = \beta^*,$$

where α^* , β^* denote the least upper bounds ($\leq \infty$) of the admissible values of the respective indices α , β .

This is quite a refined manifestation of a general principle formulated by Paul Lévy [8], pp. 264-266.

The proof will be based on (*) in a manner which will make it clear that the replacement of (*) by (* bis) leads to the following variant of (xiv):

(xiv bis) *The assertions of (xiv) remain true if (75) is replaced by*

$$(75 \text{ bis}) \quad \alpha > 0 \text{ and } f^*(t) = O(\exp -t^\alpha) \text{ as } t \rightarrow \infty,$$

and (76) by

$$(76 \text{ bis}) \quad \beta > 0 \text{ and } F^*(x) = -\frac{1}{2}f(+0)x + O(\exp -x^\beta) \text{ as } x \rightarrow +0;$$

(it being understood that α^* and β^* in (77) now refer to (75 bis) and (76 bis), respectively).

First, the assumptions of (xiv) are those of (x), if $f(x)$ is normalized by (52). Hence, (51) is applicable. Let (51) be written in the form

$$(78) \quad h(t) = \sum_{n=1}^{\infty} f^*(2\pi nt),$$

where $h(t)$ is defined by

$$(79) \quad 2h(t) = F^*(1/t) + \frac{1}{2}f(+0)/t, \quad (t > 0).$$

Next, suppose that there exists an α satisfying (75). Then (72) shows that (69) is fulfilled by the function (78). Hence, if (78) is identified with (70), where $g(t) = f^*(2\pi t)$, it follows from (*) that both (71) and (68) hold in the present case. But (71) now appears in the form

$$f^*(2\pi t) = \sum_{n=1}^{\infty} \mu(n)h(nt),$$

whereas (79) shows that (68) is identical with (76).

On the other hand, if (75) is replaced by (76) as an assumption, then (79) shows that (76) is satisfied. Hence, if (78) is identified, as before, with (70), it is seen from (79) that (68) is fulfilled by $g(t) = f^*(2\pi t)$, which means that (75) is satisfied.

This proves (xiv), since the remaining assertion, (77), is clear from (67). In addition, the last two formula lines supply the following by-product:

(xv) *If (52), (74) and either (75) or (76) are satisfied by $f(x)$, then the Möbius expansion,*

$$(80) \quad f^*(2\pi/t) = \frac{1}{2} \sum_{n=1}^{\infty} \mu(n) F^*(t/n),$$

of the Fourier cosine transform, (47), in terms of the Euler-Maclaurin function, (45), is valid for every $t > 0$.

Actually, the by-product is

$$(80') \quad f^*(2\pi/t) = \frac{1}{2} \sum_{n=1}^{\infty} \mu(n) \{F^*(t/n) + \frac{1}{2}f(+0)t/n\},$$

and so (80) follows only if either $f(+0)$ is 0 or recourse is had to

$$(81) \quad \sum_{n=1}^{\infty} \mu(n)/n = 0,$$

that is, to the prime number theorem. This means that (xv) is *elementary or just as deep as the prime number theorem according as $f(+0) = 0$ or $f(+0) \neq 0$.*

In (xv), the assumptions are the same as in (x). Under the additional assumption made in (xi), the assertion of (xv) can be dualized, as follows:

(xvi) *If (52), (74) and (28) are satisfied by $f(x)$, and if the Fourier cosine transform,*

$$(82) \quad f^*(t) = \int_0^{\infty} f(x) \cos tx \, dx = \lim_{x \rightarrow \infty} \int_0^x,$$

remains convergent at $t = 0$, then the Euler-Maclaurin function,

$$(83) \quad F^*(t) = t \sum_{n=1}^{\infty} f(nt) - \int_0^{\infty} f(x) \, dx,$$

provides a Möbius expansion,

$$(84) \quad xf(x) = \sum_{n=1}^{\infty} \mu(n) F^*(nx)/n,$$

of $f(x)$ for every $x > 0$.

What will now follow directly, namely,

$$(84') \quad xf(x) = \sum_{n=1}^{\infty} \mu(n) \{F^*(nx) + f^*(0)\}/n,$$

will reduce to (84) *without* recourse to (81) only if $f^*(0) = 0$; so that (xvi) is elementary or involves precisely the prime number theorem according as the value of the integral (53) is or is not 0.

The assumptions of (xvi) are identical with those of (xi). Hence, (56) is applicable. On the other hand, (53) shows that (83) can be written in the form

$$(78') \quad h(t) = \sum_{n=1}^{\infty} f(nt),$$

if $h(t)$ is defined by

$$(79') \quad h(t) = F^*(t)/t + f^*(0)/t.$$

The last two formula lines take over the parts played by (78) and (79) in the proof of (xiv), (xv). In fact, it is clear that the balance of the proof of (xvi) is the same as, via (*) and (xiv), the proof of (xv) was.

6. An instance of explicit interest will now be considered: $f(x) = e^{-x^\lambda}$, where $\lambda > 0$. Then, since

$$f^*(0) = \int_0^{\infty} f(x) dx = \int_0^{\infty} e^{-x^\lambda} dx = \Gamma(1 + 1/\lambda),$$

(83) becomes $F^*(t) = t \sum_{n=1}^{\infty} e^{-t^\lambda n^\lambda} - \Gamma(1 + 1/\lambda)$. This means that $F^*(t^{1/\lambda})/t^{1/\lambda}$ is the function

$$\sum_{n=1}^{\infty} e^{-tn^\lambda} - \Gamma(1 + 1/\lambda).$$

But the latter function is known to be the entire function

$$\sum_{m=0}^{\infty} (-1)^m \xi(-\lambda m) t^m / m!,$$

if $0 < \lambda < 1$; cf. Mellin [9], p. 12. Hence, (80) becomes

$$f^*(2\pi/t) = \frac{1}{2} \sum_{n=1}^{\infty} \mu(n) \sum_{m=0}^{\infty} (-1)^m \zeta(-\lambda m) (t/n)^{\lambda m+1} / m!,$$

where $0 < \lambda < 1$.

The contribution of $m = 0$ to this repeated sum is

$$\frac{1}{2} \sum_{n=1}^{\infty} \mu(n) \zeta(0) t/n = \frac{1}{2} t \zeta(0) \sum_{n=1}^{\infty} \mu(n)/n = 0,$$

by (81). Hence,

$$f^*(2\pi/t) = \frac{1}{2} t \sum_{n=1}^{\infty} \mu(n) \sum_{m=0}^{\infty} (-t^\lambda)^m \zeta(-\lambda m) / (m! n^{\lambda m+1}).$$

This result is independent of the prime number theorem. For, on the one hand, (81) has been used twice (first in the reduction of (80') to (80) and then in the omission of the summation value $m = 0$) and, on the other hand, it is clear that these two steps can be united into one which avoids (81) entirely.

Since $\mu(n) = O(1)$, it is easily seen (from the behavior of $\zeta(-x)$ for large positive x) that the repeated sum on the right of the last formula line actually is an absolutely convergent double series; in fact, as mentioned before, Mellin's power series converges for arbitrarily large t , since $0 < \lambda < 1$. Hence, the order of summations can be interchanged. The contribution of what then becomes the interior summation is

$$\sum_{n=1}^{\infty} \mu(n) (-t^\lambda)^m \zeta(-\lambda m) / (m! n^{\lambda m+1}) = (-t^\lambda)^m / m! \zeta(-\lambda m) / \zeta(\lambda m + 1),$$

since $\sum_{n=1}^{\infty} \mu(n)/n^s = \zeta(s)$ when $s > 1$. Accordingly,

$$f^*(2\pi/t) = \frac{1}{2} t \sum_{n=1}^{\infty} (-t^\lambda)^n / n! \zeta(-\lambda n) / \zeta(\lambda n + 1).$$

But $f(x)$ was the function e^{-x^λ} ; so that $f^*(t) = \int_0^\infty e^{-x^\lambda} \cos tx \, dx$, by (82).

Consequently,

$$\int_0^\infty e^{-x^\lambda} \cos 2\pi tx \, dx = \frac{1}{2} t^{-1} \sum_{n=1}^{\infty} \zeta(-\lambda n) / \zeta(\lambda n + 1) (-t^\lambda)^n / n!.$$

The coefficients of this power series can be reduced by using Riemann's functional equation,

$$\frac{1}{2} \zeta(1-s) = (2\pi)^{-s} \zeta(s) \Gamma(s) \cos \frac{1}{2} \pi s.$$

For $s = 1 + \lambda n$, this gives

$$\frac{1}{2}\xi(-\lambda n)/\xi(1+\lambda n) = (2\pi)^{-1}(2\pi)^{-\lambda n}\Gamma(1+\lambda n)(-\sin \frac{1}{2}\pi\lambda n).$$

Hence, the preceding representation of the Fourier cosine integral $f^*(2\pi t)$ can be written in the form

$$-(2\pi t)^{-1} \sum_{n=1}^{\infty} \Gamma(1+\lambda n) \sin \frac{1}{2}\pi\lambda n (-t^{-\lambda})^n (2\pi)^{-\lambda n}/n!.$$

Accordingly, the final result is as follows:

If $0 < \lambda < 1$, then

$$\int_0^{\infty} e^{-x\lambda} \cos tx \, dx = \sum_{n=1}^{\infty} (-1)^{n+1} \sin \frac{1}{2}\pi\lambda n \Gamma(\lambda n + 1)/n! \, t^{-\lambda n-1},$$

where t is arbitrary ($\neq 0$). Cf. [12].

In the limiting case $\lambda = 1$, this expansion is equivalent to that of $(1+t^{-2})^{-1}$, a geometric progression valid for certain, but not for all, values of t . It is easy to see from the above proof that, if $\lambda > 1$, the power series, which then diverges everywhere, is an asymptotic expansion (as $t \rightarrow \infty$) of the function $f^*(t)$. It should be noted that, if λ is an even integer, then all coefficients vanish, and so

$$f^*(t) = O(t^{-N}) \text{ as } t \rightarrow \infty,$$

where N is arbitrarily large.

A final remark is needed for the justification of the above application of (80), since (xv) assumes either (75) or (76). But (75), with $\alpha = 1 + \lambda$, is clear from the first term of the final expansion, which can, of course, be verified directly, and also (76) can be verified directly, by using the indications of Lévy, referred to after (xiv); in fact, Lévy formulates his general principle precisely in connection with the above $f^*(t)$, the Fourier transform of a symmetric stable distribution (cf. [8], pp. 264-277).

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES

- [1] K. Ananda-Rao, "On Lambert's series," *Proceedings of the London Mathematical Society*, ser. 2, vol. 19 (1921), pp. 1-20.
- [2] T. J. I'A. Bromwich and G. H. Hardy, "The definition of an infinite integral as the limit of a finite or infinite series," *Quarterly Journal of Mathematics*, vol. 39 (1908), pp. 222-240.

- [3] H. Hahn, "Ueber eine Verallgemeinerung der Fourierschen Integralformel," *Acta Mathematica*, vol. 49 (1926), pp. 301-353.
- [4] G. H. Hardy, "Theorems connected with MacLaurin's test for the convergence of series," *Proceedings of the London Mathematical Society*, ser. 2, vol. 9 (1911), pp. 126-144.
- [5] ——— and J. E. Littlewood, "On a Tauberian theorem for Lambert's series, and some fundamental theorems in the analytic theory of numbers," *ibid.*, vol. 19 (1921), pp. 21-29.
- [6] ———, "Notes on the theory of series (XX): On Lambert series," *ibid.*, vol. 41 (1936), pp. 257-270.
- [7] H. Lebesgue, *Leçons sur les séries trigonométriques*, Paris, 1906.
- [8] P. Lévy, *Calcul des probabilités*, Paris, 1925.
- [9] H. Mellin, "Die Dirichletschen Reihen, die zahlentheoretischen Funktionen und die unendlichen Produkte von endlichem Geschlecht," *Acta Societatis Scientiarum Fennicae*, vol. 31 (1903), no. 2.
- [10] A. Pringsheim, "Ueber neue Gültigkeitsbedingungen für die Fouriersche Integralformel," *Mathematische Annalen*, vol. 68 (1910), pp. 367-408; Nachtrag, *ibid.*, vol. 71 (1912), pp. 289-298.
- [11] A. Wintner, *An Arithmetical Approach to Ordinary Fourier Series*, Baltimore, 1945.
- [12] ———, "The singularities of Cauchy's distributions," *Duke Mathematical Journal*, vol. 8 (1941), pp. 678-681.

APPLICATIONS OF INDUCED CHARACTERS.*

By RICHARD BRAUER.

1. Introduction. In a previous paper,¹ the following theorem on induced characters of groups was proved: If \mathfrak{G} is a group of finite order g , every character χ of \mathfrak{G} can be written in the form $\chi = \sum a_p \omega_p^*$ where the a_p are rational integers and where the ω_p^* are characters of \mathfrak{G} induced by linear characters ω_p of subgroups \mathfrak{S}_p of \mathfrak{G} . If we call a group *elementary*, if it is a direct product $\mathfrak{U} \times \mathfrak{B}$ of a group \mathfrak{U} of prime power order and a cyclic group \mathfrak{B} of an order prime to the order of \mathfrak{U} , then we may assume that all the groups \mathfrak{S}_p are elementary groups. This can be seen at once from the proof of the theorem.

As an immediate consequence of this result, it will be shown in **2** that every representation of a group \mathfrak{G} of order g can be written in the field of the g -th roots of unity. Our new approach to this problem is simpler and more elementary than that given in an earlier investigation.² At the same time it yields stronger results. For instance, if n is the least common multiple of the orders of the elements of \mathfrak{G} , then every representation of \mathfrak{G} can be written in the field of the n -th roots of unity.

3 deals with a method of determining the irreducible characters of a group of finite order. If ω is a character of a subgroup \mathfrak{S} of \mathfrak{G} , we need the following information in order to be able to construct the character ω^* of \mathfrak{G} . We have to know (a) the number k of classes $\mathfrak{R}_1, \mathfrak{R}_2, \dots, \mathfrak{R}_k$ of conjugate elements in \mathfrak{G} and the number g_i of elements in \mathfrak{R}_i ; (b) the number l of classes $\mathfrak{Q}_1, \mathfrak{Q}_2, \dots, \mathfrak{Q}_l$ of conjugate elements in \mathfrak{S} and the number h_j of elements in \mathfrak{Q}_j ; (c) the value $i = i(j)$ such that $\mathfrak{Q}_j \subseteq \mathfrak{R}_{i(j)}$, $j = 1, 2, \dots, l$; (d) the value $\omega(\mathfrak{Q}_j)$ of ω for \mathfrak{Q}_j .³ If this information is given for all elementary subgroups \mathfrak{S} of \mathfrak{G} and all linear characters ω of \mathfrak{S} , then it is shown that all irreducible characters of \mathfrak{G} can be constructed. In order to obtain all linear characters ω of \mathfrak{S} , we have to know the normal subgroups \mathfrak{S}_0 of \mathfrak{S} with

* Received February 9, 1947.

¹ R. Brauer, *Annals of Mathematics*, vol. 48 (1947), pp. 502-514.

² R. Brauer, *American Journal of Mathematics*, vol. 67 (1945), pp. 461-471.

³ If ω is a character of \mathfrak{S} , H an element of the class \mathfrak{Q} of \mathfrak{S} , we denote by $\omega(\mathfrak{Q})$ the value $\omega(H)$ taken by ω for the element H .

cyclic factor group $\mathfrak{G}/\mathfrak{G}_0 = \{\mathfrak{G}_0 F\}$ and we have to know the exponent $\rho = \rho(j)$ for which the class \mathfrak{A}_j belongs to the coset $\mathfrak{G}_0 F^\rho$, $j = 1, 2, \dots, l$. It seems remarkable that it is possible to construct the characters of \mathfrak{G} on the basis of so little information concerning the structure of \mathfrak{G} .

In 4, an analogue of the theorem on induced characters for the case of modular characters is given. If p^a is the highest power of the prime p dividing the group order g , then as an application the number N of ordinary irreducible representations \mathfrak{B} of \mathfrak{G} is determined whose degree z is divisible by p^a . The number N is obtained as the number of representations of 1 by means of a quadratic form with integral rational coefficients.

2. Representation of groups in cyclotomic fields. We first prove:

THEOREM 1. *Let \mathfrak{G} be a group of finite order and let n be the least common multiple of the orders of the elements of \mathfrak{G} . Every representation of \mathfrak{G} can be written in the field of the n -th roots of unity.*

Proof. It is sufficient to prove the theorem for irreducible representations \mathfrak{B} of \mathfrak{G} . The field K of the n -th roots of unity certainly contains the character χ of \mathfrak{B} as well as all characters ω of subgroups \mathfrak{G} of \mathfrak{G} . In particular, every linear representation \mathfrak{M} of a subgroup \mathfrak{G} lies in K , and the same holds for the representation \mathfrak{M}^* of \mathfrak{G} induced by \mathfrak{M} . According to the theorem on induced characters quoted in the introduction, we have

$$(1) \quad \chi = \sum a_\rho \omega_\rho^*$$

where the a_ρ are rational integers and where the ω_ρ^* are characters induced by linear characters ω_ρ of subgroups \mathfrak{G}_ρ of \mathfrak{G} . If p is a fixed prime number, then (1) shows that there exists at least one ω_ρ^* which contains χ with a multiplicity t prime to p . The representation \mathfrak{M}^* belonging to ω_ρ^* lies in the field K and contains \mathfrak{B} with the multiplicity t . This implies⁴ that the Schur index m of \mathfrak{B} with respect to K divides t and is, therefore, prime to p . Since this holds for every prime p , we have $m = 1$. Then, \mathfrak{B} can be written in the field K as was to be shown.

The same procedure yields a slightly better result. Let K_0 be the field $P(\chi)$ obtained by adjunction of the character χ of \mathfrak{B} to the field P of rational numbers, and let m_0 be the Schur index of \mathfrak{B} with respect to K_0 . As was shown by Schur,⁵ m_0 divides the degree z of \mathfrak{B} . If p is a prime factor of m_0 , we wish to adjoin an α -th root of unity to K_0 such that the Schur index of \mathfrak{B}

⁴ See I. Schur, *Sitzungsberichte Preuss. Akad. Wiss.* (1906), pp. 164-184.

⁵ *Loc. cit.*⁴

with respect to the extended field is no longer divisible by p . Choose as above ω_p^* in (1) such that ω_p^* contains χ with a multiplicity t not divisible by p . If the values of ω_p lie in the field of the α -th roots of unity, then the adjunction of the α -th roots of unity to K_0 will have the desired effect. However, this adjunction is equivalent to the simultaneous adjunction of roots of unity of prime power exponents q^b where all q^b divide α . If $q \not\equiv 0, 1 \pmod{p}$, the degree of a (q^b) -th root of unity with respect to K_0 is not divisible by p . Hence the adjunction of the (q^b) -th roots of unity cannot change the power of p in the Schur index in this case. Similarly, if $q \neq p$, $b > 1$, the adjunction of the (q^b) -th roots of unity can be replaced by adjunction of the q -th roots of unity and the same reduction of the power of p in the Schur index will be achieved. Finally, if $q = p$, $b = 1$, the adjunction of the corresponding roots of unity is again superfluous. We thus see that we may replace α by a divisor β which contains only prime factors q of the form $q \equiv 0, 1 \pmod{p}$, the factors $q \equiv 1$ all with the exponent 1, and the factor $q = p$ with an exponent $b \neq 1$ (possibly $b = 0$). Since we may assume that \mathfrak{G} contains elements of the order α , it will also contain elements of order β .

If this procedure is applied for all prime divisors of m_0 , the following theorem is obtained.

THEOREM 2. *Let p_1, p_2, \dots, p_r be the distinct primes which divide the Schur index of the irreducible representation \mathfrak{Z} of \mathfrak{G} with respect to the field of the character of \mathfrak{Z} . (Then the p_i divide the degree of \mathfrak{Z} .) We can find a system of elements G_1, G_2, \dots, G_r of \mathfrak{G} with the following properties:*

1. *The order β_p of G_p contains only primes $q \equiv 0, 1 \pmod{p_p}$. If the prime p_p appears in β_p , it appears with an exponent ≥ 2 . All other prime factors of β_p appear only with the exponent 1.*

2. *If v is the least common multiple of $\beta_1, \beta_2, \dots, \beta_r$ then \mathfrak{Z} can be written in the field which is obtained from the field of the character of \mathfrak{Z} by an adjunction of a v -th root of unity.*

We add a remark for the case that the character χ of the representation \mathfrak{Z} is real. If the degree z of \mathfrak{Z} is odd, then A. Speiser⁶ showed that \mathfrak{Z} can be written in the field K_0 of the character χ . If z is even, the Schur index m_0 divides 2.⁷ It is then sufficient to consider only the prime 2 in Theorem 2.

⁶ A. Speiser, *Mathematische Zeitschrift*, vol. 5 (1919), pp. 1-6.

⁷ R. Brauer, *Sitzungsberichte Preuss. Akad. Wiss.* (1926), pp. 410-416; R. Brauer, H. Hasse, and E. Noether, *Journal für die reine und angewandte Mathematik*, vol 167

As is easily seen, the number β_1 can be taken here either as an odd prime or as a power of 2. This gives

THEOREM 3. *If \mathfrak{B} is an irreducible representation with a real character, there exists an element G in \mathfrak{G} whose order β is either an odd prime or a power of 2, such that \mathfrak{B} can be written in the field obtained by adjunction of the β -th roots of unity to the field of characters.*

3. Construction of the characters of a group \mathfrak{G} of finite order.

If \mathfrak{H} is a subgroup of order h of \mathfrak{G} and if ω is a character of \mathfrak{H} , the induced character ω^* of \mathfrak{G} is given by

$$(2) \quad \omega^*(G) = (1/h) \sum \omega(RGR^{-1})$$

where R on the right ranges over all g elements of \mathfrak{G} , and where $\omega(X) = 0$ if X is not an element of \mathfrak{H} . If G belongs to the class \mathfrak{R}_i of \mathfrak{G} , only terms $\omega(\mathfrak{L}_j)$ will appear on the right side of (2) for which \mathfrak{L}_j is a class of \mathfrak{H} which is contained in \mathfrak{R}_i . If \mathfrak{R}_i contains g_i elements and \mathfrak{L}_j contains h_j elements, the term $\omega(\mathfrak{L}_j)$ appears with the multiplicity gh_j/g_i and (2) may be written in the form

$$(3) \quad \omega^*(\mathfrak{R}_i) = (g/hg_i) \sum h_j \omega(\mathfrak{L}_j),$$

the sum extending over all classes \mathfrak{L}_j of \mathfrak{H} with $\mathfrak{L}_j \subseteq \mathfrak{R}_i$.

It is now evident that if the information (a), (b), (c), (d) mentioned in the introduction is given, the character ω^* as a function of \mathfrak{R}_i is completely determined. Apply this for all elementary subgroups \mathfrak{H} of \mathfrak{G} and for all linear characters ω of \mathfrak{H} . In this manner, we obtain a system of characters $\omega_1^*, \omega_2^*, \dots, \omega_r^*$ of \mathfrak{G} such that every character of \mathfrak{G} can be written in the form $\chi = \sum a_p \omega_p^*$ with integral rational coefficients.

Conversely, every ω_p^* can be written as a linear combination of the irreducible characters $\chi_1, \chi_2, \dots, \chi_k$ of \mathfrak{G} with integral rational coefficients. The same holds for a linear combination $\xi = \sum x_p \omega_p^*$ with integral rational coefficients x_p , say

$$(4) \quad \xi = \sum_{p=1}^r x_p \omega_p^* = \sum_{\kappa=1}^k u_\kappa \chi_\kappa.$$

The orthogonality relations for group characters yield

$$(1/g) \sum_i g_i \xi(\mathfrak{R}_i) \bar{\xi}(\mathfrak{R}_i) = (1/g) \sum_{\rho, \sigma} x_\rho x_\sigma \sum_i g_i \omega_\rho^*(\mathfrak{R}_i) \bar{\omega}_\sigma^*(\mathfrak{R}_i) = \sum_k u_k^2.$$

Set

(1932), pp. 399-404. In the first of these papers, it was shown that the exponent of the representation is 2, and in the second paper that the exponent is equal to the Schur index.

$$(5) \quad m_{\rho\sigma} = (1/g) \sum g_i \omega_{\rho}^*(\mathfrak{R}_i) \bar{\omega}_{\sigma}(\mathfrak{R}_i).$$

Then $m_{\rho\sigma}$ is a non-negative rational integer and we have

$$(6) \quad \sum_{\rho, \sigma} x_{\rho} x_{\sigma} m_{\rho\sigma} = \sum_{\kappa} u_{\kappa}^2.$$

In particular, if the expression (6) is equal to 1, then either ξ or $-\xi$ is an irreducible character χ_{κ} of \mathfrak{G} . It is easy to decide which of these two cases we have. Indeed, let \mathfrak{R}_1 be the class which contains the 1-element of \mathfrak{G} . If $\xi(\mathfrak{R}_1) > 0$, then ξ itself is an irreducible character while in the other case $-\xi$ is an irreducible character.⁸

In order to find all irreducible characters of \mathfrak{G} , we have to find all solutions of the Diophantine equation

$$(7) \quad \sum x_{\rho} x_{\sigma} m_{\rho\sigma} = 1$$

in rational integers x_{ρ} . The coefficients $m_{\rho\sigma}$ of the quadratic form on the left side can be found, if the ω_{ρ}^* are known. Only solutions x_{ρ} are to be used for which $\sum x_{\rho} \omega_{\rho}^*(\mathfrak{R}_1) > 0$. There are exactly k distinct expressions $\xi = \sum x_{\rho} \omega_{\rho}^*$ formed by means of such solutions x_{ρ} . These k expressions are the k irreducible characters of \mathfrak{G} .

THEOREM 4. *Suppose that the number k of classes $\mathfrak{R}_1, \mathfrak{R}_2, \dots, \mathfrak{R}_k$ of conjugate elements of \mathfrak{G} and the number g_i of elements in \mathfrak{R}_i are known. Suppose that a complete system of elementary subgroups \mathfrak{S} of \mathfrak{G} is given, (subgroups conjugate in \mathfrak{G} may be considered as not essentially different). Assume further that for each \mathfrak{S} the number l of classes $\mathfrak{L}_1, \mathfrak{L}_2, \dots, \mathfrak{L}_l$ of conjugate elements of \mathfrak{S} and the number h_j of elements of \mathfrak{L}_j is known, that it is known to which class \mathfrak{R}_i the elements of \mathfrak{L}_j belong and that the values $\omega(\mathfrak{L}_j)$ of the linear characters ω of \mathfrak{S} are known. Then the irreducible characters of \mathfrak{G} are completely determined.*

As already remarked in the introduction, the construction of all linear characters $\omega(\mathfrak{L}_j)$ of \mathfrak{S} requires only the knowledge of all normal subgroups \mathfrak{S}_0 with cyclic factor group of the elementary group \mathfrak{S} and the information to which particular coset (mod \mathfrak{S}_0) the class \mathfrak{L}_j belongs.

Since the characters $\omega_1^*, \omega_2^*, \dots, \omega_r^*$ are, in general, linearly dependent, the equation (7) has, in general, infinitely many solutions. However, it can be seen without difficulty that the solutions can be found in a finite number of steps.

⁸ This method has been used by I. Schur in order to find the characters of special groups.

4. An analogue for modular characters. Let p now be a fixed prime number. We prove

THEOREM 5. *If Φ is the character of an indecomposable constituent of the modular regular representation of $\mathfrak{G} \pmod{p}$, then Φ can be written in the form*

$$\Phi = \sum_{\sigma} a_{\sigma} \omega_{\sigma}^*$$

where the a_{σ} are rational integers and where the ω_{σ}^* are characters of \mathfrak{G} induced by linear characters ω_{σ} of elementary subgroups \mathfrak{S}_{σ} of orders prime to p .

Proof. We first observe that Φ may be considered as an ordinary (reducible or irreducible) character of \mathfrak{G} which vanishes for the p -singular classes of \mathfrak{G} .⁹ All characters ω_{σ}^* in Theorem 5 vanish for the same classes.

As in the proof of the theorem on induced characters (see ¹), it is sufficient to show that if a congruence

$$(8) \quad \sum c_i \omega_{\sigma}^*(\mathfrak{R}_i) \equiv 0 \pmod{q^r}$$

modulo a prime ideal power q^r of a suitable algebraic number field has q -integral coefficients c_i and holds for all characters ω_{σ}^* in Theorem 5, then the corresponding congruence

$$(9) \quad \sum c_i \Phi(\mathfrak{R}_i) \equiv 0 \pmod{q^t}$$

holds for Φ . It is sufficient to restrict the summation to p -regular classes \mathfrak{R}_i .

If q does not divide p , it follows at once from Theorem 2 of the paper quoted in ¹ that (8) implies (9). It remains to treat the case that q is a prime ideal divisor of p . Let A be an element of \mathfrak{R}_i and let $\xi_0, \xi_1, \dots, \xi_{a-1}$ denote the linear characters of the cyclic group $\{A\}$. Set

$$\psi(A^v) = \sum_{i=0}^{a-1} \xi_i(A) \xi_i(A^v).$$

Then

$$(10) \quad \psi(A^v) = \begin{cases} \alpha, & A^v = A \\ 0, & A^v \neq A \end{cases}$$

The induced expression is $\psi^*(G) = (1/\alpha) \sum \psi(RGR^{-1})$ where R ranges over all elements of \mathfrak{G} . Now, (10) yields

$$\psi^*(G) = \begin{cases} n(A), & G \text{ in } \mathfrak{R}_i \\ 0, & G \text{ not in } \mathfrak{R}_i \end{cases}$$

⁹ Cf. R. Brauer and C. Nesbitt, *Annals of Mathematics*, vol. 42 (1942), pp. 556-590, in particular, equation (9) and the argument in § 14.

where $n(A) = g/g_i$ is the order of the normalizer of A in \mathfrak{G} . Substituting this for ω_{σ}^* in (8), we find

$$c_i n(A) \equiv 0 \pmod{q^t}.$$

Now $\Phi(\mathfrak{R}_i)$ is divisible by the highest power of q dividing $n(A)$. Hence

$$c_i \Phi(\mathfrak{R}_i) \equiv 0 \pmod{q^t}.$$

This implies (9), and Theorem 5 is proved.

The linear combinations of the characters $\omega_1^*, \omega_2^*, \dots, \omega_s^*$ with integral rational coefficients form a module Ω . If elements of Ω are linearly dependent, there exists a linear relation with integral rational coefficients. This is seen at once when the elements of Ω are expressed by the irreducible characters of \mathfrak{G} . Let $\psi_1, \psi_2, \dots, \psi_w$ be a basis of Ω . Then $\psi_1, \psi_2, \dots, \psi_w$ are linearly independent in the field of all numbers.

In particular, the characters Φ_1, Φ_2, \dots of the distinct indecomposable constituents of the modular regular representation of \mathfrak{G} belong to Ω and they are linearly independent. Every element of Ω vanishes for all p -singular classes \mathfrak{R}_i of \mathfrak{G} and can, therefore, be expressed by the Φ_i with integral rational coefficients.¹⁰ Hence the Φ_i also form a basis of Ω . This shows that the number w of basis elements of Ω is equal to the number of distinct Φ_i , that is, to the number of p -regular classes \mathfrak{R}_i in \mathfrak{G} . Further, the ψ_i and the Φ_i are connected by a unimodular linear transformation with integral rational coefficients,

$$(11) \quad \psi_i = \sum b_{ij} \Phi_j.$$

While it is of course possible to determine a basis $\psi_1, \psi_2, \dots, \psi_w$ of Ω when the ω_{σ}^* are known, it seems that the Φ_j themselves cannot always be determined on the basis of this information.

It follows from (11) and the orthogonality relations for modular group characters that

$$(12) \quad q_{\alpha\beta} = (1/g) \sum_i g_i \psi_{\alpha}(\mathfrak{R}_i) \bar{\psi}_{\beta}(\mathfrak{R}_i) = \sum_{\rho, \sigma} b_{\alpha\rho} c_{\rho\sigma} b_{\beta\sigma}$$

¹⁰ If a linear combination ξ with integral rational coefficients of the ordinary characters χ_1, χ_2, \dots of \mathfrak{G} vanishes for all p -singular elements of \mathfrak{G} , then a consideration of ranks shows that ξ can be written in the form $\xi = \sum h_i \Phi_i$ with complex coefficients h_i . The orthogonality relations for modular group characters yield $h_i = (1/g) \sum \xi(R^{-1}) \phi_i(R)$ where ϕ_1, ϕ_2, \dots are the modular irreducible characters of \mathfrak{G} and where R ranges over all p -regular elements of \mathfrak{G} . Each $\phi_i(R)$ can be written as a linear combination of the $\chi_j(R)$ with integral rational coefficients. Substituting this expression for $\phi_i(R)$, we easily see that the h_i are rational integers.

where the $c_{p\sigma}$ are the Cartan invariants of \mathfrak{G} . The matrix with the coefficients $q_{\alpha\beta}$ is equal to BCB' where $B = (b_{\alpha\beta})$, $C = (c_{\alpha\beta})$. The corresponding quadratic form is equivalent to the form with the matrix C . This yields

THEOREM 6. *If the characters ω_{σ}^* in Theorem 5 are known, a quadratic form can be found which is equivalent to the form whose matrix is the Cartan matrix of \mathfrak{G} for p .*

Consider an element ξ of Ω ,

$$\xi = \sum x_{\sigma} \psi_{\sigma}.$$

It follows from (12) that

$$(13) \quad (1/g) \sum g_i \xi(\mathfrak{R}_i) \bar{\xi}(\mathfrak{R}_i) = \sum_{\rho, \sigma} x_{\rho} x_{\sigma} q_{\rho\sigma}.$$

If the expression (13) is equal to 1, then $\pm \xi$ is an irreducible ordinary character of \mathfrak{G} . Since ξ vanishes for all p -singular classes of \mathfrak{G} , its degree is divisible by the highest power p^a of p which divides g .¹¹ Conversely, if ξ is an irreducible ordinary character of \mathfrak{G} whose degree is divisible by p^a , then ξ vanishes for p -singular classes and belongs, therefore, to Ω . If we set $\xi = \sum x_{\sigma} \psi_{\sigma}$, the coefficients x_{σ} give a solution of

$$\sum x_{\rho} x_{\sigma} q_{\rho\sigma} = 1.$$

We thus have

THEOREM 7. *Let p^a be the highest power of p which divides the order g of \mathfrak{G} . The number of ordinary irreducible representations of \mathfrak{G} whose degree is divisible by p^a is equal to the number of representations of 1 by the quadratic form in Theorem 6. (We count x_1, x_2, \dots, x_w and $-x_1, -x_2, \dots, -x_w$ as the same representation.)*

The number determined in Theorem 7 can also be characterized as the number of blocks of defect 0 of \mathfrak{G} (for p). To some extent, Theorem 7 fills a gap left in the investigation of the blocks of a given group.¹²

UNIVERSITY OF TORONTO.

¹¹ See the paper quoted in ⁹.

¹² R. Brauer, *Proceedings of the National Academy of Sciences*, vol. 30 (1944), pp. 109-114, vol. 32 (1946), pp. 182-186 and 215-219.

PRINCIPAL SOLUTIONS OF DIFFERENCE EQUATIONS.*

By WALTER STRODT.¹

PART I. Introduction.

One of the central problems in the theory of difference equations is that of eradicating, as far as possible, the arbitrary elements in the manifold of solutions.² In brief, a difference equation has so *many* solutions, (in those cases where the complete manifold has been determined, involving one or more arbitrary periodic functions), that the property of a function of being a solution of a given difference equation is of little value in the investigation of the behavior of the function. In order that a function may be studied by means of a difference equation which it satisfies, it is essential in many cases that the function be singled out as playing a somehow distinguished role in the manifold of solutions of that equation.

The concept of principal solution introduced by N. E. Nörlund³ serves, in the cases of those equations for which the concept has been defined, to distinguish a unique solution, or a unique several-parameter family of solutions. Usually the solutions so distinguished are the most interesting solutions of the equation, being, roughly speaking, the solutions of minimal rate of growth at infinity.⁴

The core of Nörlund's work on principal solutions is the investigation of the two types of equation

$$(1) \quad f(x + \omega) - f(x) = \omega \phi(x),$$

for which the principal solution is defined⁵ to be the solution given by the formula

* Received February 7, 1947.

¹ National Research Fellow, at Harvard University; on leave from Columbia University.

² Cf. R. D. Carmichael, "The present state of the difference calculus and the prospect for the future," *American Mathematical Monthly*, vol. 31 (1924), pp. 172 ff.

³ N. E. Nörlund, *Vorlesungen über Differenzenrechnung*, Berlin, 1924, Chapters III and IV.

⁴ The concept of minimal rate of growth has never been made sufficiently precise to serve as a *definition* of principal solution. It will be noted that in both Nörlund's definition, and the author's given below, appeal is made to other considerations.

⁵ Nörlund, *loc. cit.*, pp. 40-43.

$$(2) \quad f(x) = \lim_{\eta \rightarrow 0} \left[\int_a^\infty \phi(z) e^{-\eta \lambda(z)} dz - \omega \sum_{s=0}^\infty \phi(x + s\omega) e^{-\eta \lambda(x+s\omega)} \right] \quad \dagger$$

(where a is an arbitrary constant, and $\lambda(z)$ is a function of the form $x^p (\log x)^q$, ($p \geq 1$, $q \geq 0$)), and

$$(3) \quad f(x + \omega) + f(x) = 2\phi(x)$$

for which the principal solution is defined to be the solution given by the formula

$$(4) \quad f(x) = \lim_{\eta \rightarrow 0} \left[2 \sum_{s=0}^\infty (-1)^s \phi(x + s\omega) e^{-\eta \lambda(x+s\omega)} \right].$$

Carmichael has remarked⁶ that in the related problem of *q-difference* equations the elimination of the arbitrary elements is a comparatively easy matter. Where for *difference* equations the crucial question seems to be asymptotic behavior at infinity, and this is difficult to make precise in a general situation, for *q-difference* equations the simple property of analyticity at $x = 0$, (or at $x = \infty$), serves to characterize the distinguished solutions.

For the purpose of studying difference equations we shall find it useful to generalize this concept of distinguished solutions of *q-difference* equations. We define a *special solution* of a *q-difference* equation to be a solution expandable in ascending, *not necessarily integral*, powers of x .⁷

Using this concept of *special solution* of *q-difference* equations we give in this paper a new definition of principal solution, for difference equations with analytic coefficients. In brief, the principal solution of a difference equation is in this paper defined as a solution embedded in *special solutions* of certain *q-difference* equations which formally approximate the given difference equation. This definition is directly applicable to both linear and non-linear equations, and is effective in singling out those solutions of the difference equation which are of minimal rate of growth at infinity. It affords a vast generalization of Nörlund's results, (in the case of analytic coefficients), and serves to link the somewhat disparate definitions (2) and (4) of Nörlund.

The concepts used below are, in part, modifications of concepts introduced by the author in a paper on non-linear difference equations.⁸

⁶ Carmichael, *loc. cit.*, p. 173.

⁷ Cf. Part II, Section C, below.

⁸ "Analytic solutions of non-linear difference equations," *Annals of Mathematics*, vol. 44 (1943), pp. 375-396.

PART II. Definitions and Notations.

A. *Quasi-neighborhoods of a point.* Let b be any non-zero number. By the *quasi-neighborhood of b , of radius R , and argument γ* , will be meant the set \mathcal{S} consisting of all points x such that $0 < |x - b| < R$, and $\arg(x - b) \neq \gamma$. In other words, a quasi-neighborhood is a circular neighborhood cut along a radius.

B. *Generalized power-series.* Let b be any complex number. By a *generalized power-series at b* is meant a series of the form

$$\sum_{k=0}^{\infty} c_k (x - b)^{\lambda_k}$$

where $0 \leq \lambda_k < \lambda_{k+1}$, ($k = 0, 1, \dots$), and $\lim_{k \rightarrow \infty} \lambda_k = \infty$, the series being convergent throughout a quasi-neighborhood of b . (We understand by $(x - b)^{\lambda_k}$ the function $e^{\lambda_k \log(x - b)}$, where in the quasi-neighborhood $0 < |x - b| < R$, $\arg(x - b) \neq \gamma$, the branch of the logarithm is to be the one such that $\gamma - \pi < \Re(\log(x - b)) \leq \gamma + \pi$.)

C. *Special solutions of q -difference equations.* Let $T_1 x, \dots, T_n x$ be analytic functions of the complex variable x , such that for some (finite) complex number b the "fixed-point equations" $T_k b = b$, ($k = 1, \dots, n$), are all valid. Let $f(x, y_1, \dots, y_n)$ be a polynomial in the indeterminates y_k with coefficients analytic at $x = b$. By a *special solution $y(x)$* of the functional equation

$$(5) \quad f(x, y(T_1 x), \dots, y(T_n x)) = 0$$

is meant any generalized power-series at b which satisfies (5) in a quasi-neighborhood of b , or an analytic continuation of such a generalized power-series, the continuation being along a radius $\arg(x - b) = \text{constant}$.

Included in this definition is the definition of the special solution of a q -difference equation, in which $T_k x = q_k x$ for some constants q_k , ($k = 1, \dots, n$), and in which $b = 0$.

D. *$Q(\alpha)$ -sequences.* Let q_1, q_2, \dots be a sequence of complex numbers, none of which equals unity, the limit of the sequence being unity. Let b_1, b_2, \dots be a sequence of complex numbers, such that $\lim_{t \rightarrow \infty} b_t(1 - q_t) = 1$. If there exists a number α with $-\pi < \alpha \leq \pi$ such that $\arg b_t - \alpha = O(|b_t|^{-1})$, (where O is the Landau order symbol), we shall say that the sequence (q_t, b_t) , ($t = 1, 2, \dots$), is a $Q(\alpha)$ -sequence.

E. *Approximating q -difference equations.* Let

$$(6) \quad f(x, y(x + \omega_1), \dots, y(x + \omega_n)) = 0$$

be a difference equation, in which the ω_k are given complex numbers, and $f(x, y_1, \dots, y_n)$ is a polynomial in the indeterminates y_k with coefficients analytic in a region \mathcal{D} which has the property that whenever x_0 is in \mathcal{D} , so also are all the points $x_0 + \omega_1, \dots, x_0 + \omega_n$. Let (q_t, b_t) , $(t = 1, 2, \dots)$, be a $Q(\alpha)$ -sequence for some α . Let $q_{tj} = (q_t)^{\omega_j}$, where by $(q_t)^{\omega_j}$ is meant $e^{\omega_j \text{Log } q_t}$.⁹ Let $V_{tj} = b_t(1 - q_{tj})$, $(j = 1, \dots, n; t = 1, 2, \dots)$. Then the sequence of functional equations

$$(7) \quad f(x, y(q_{t1}x + V_{t1}), \dots, y(q_{tn}x + V_{tn})) = 0,$$

$(t = 1, 2, \dots)$, will be called an *approximating sequence of q -difference equations*.¹⁰ Any equation in the sequence (7) will be referred to as an *approximating q -difference equation*.

F. *Primary solutions of difference equations.* Given the difference equation (6), and a function $y(x)$ which is analytic in a region \mathcal{I} having the property that whenever x is in \mathcal{I} , so also are all points $x + \omega_j$, $(j = 1, \dots, n)$. Let α be a number such that $-\pi < \alpha \leq \pi$. We shall say that $y(x)$ is a *primary solution of (6), for the region \mathcal{I} , in the direction α* , (briefly, a $Q(\alpha, \mathcal{I})$ solution of (6)), if there exists a $Q(\alpha)$ -sequence (q_t, b_t) , $(t = 1, 2, \dots)$, such that if $y_t(x)$, $(t = 1, 2, \dots)$, is chosen suitably as a special solution of the approximating q -difference equation (7), then the sequence $y_t(x)$, $(t = 1, 2, \dots)$, converges to $y(x)$ in \mathcal{I} , uniformly in every closed bounded subset of \mathcal{I} . (As is customary, we allow in this definition that at each point x of \mathcal{I} finitely many of the functions $y_t(x)$ may fail to be defined.)

G. *Principal solutions of difference equations.* Given the difference equation (6), and a function $y(x)$ which is analytic in a region \mathcal{U} such that whenever x is in \mathcal{U} so also are all the points $x + \omega_1, \dots, x + \omega_n$. We shall say that $y(x)$ is a *principal solution of (6), for the region \mathcal{U} , in the direction α* , (briefly, a $P(\alpha, \mathcal{U})$ solution of (6)), if for every closed bounded

⁹ Throughout this paper we shall use the notation $\text{Log } z$, with upper case L , to indicate that determination of $\log z$ such that $-\pi < \Re(\text{Log } z) \leq \pi$. Similarly, we shall use the notation $\text{Arg } z$, with upper case A , to indicate that determination of $\arg z$ such that $-\pi < \text{Arg } z \leq \pi$.

¹⁰ It is readily verified that limit $V_{tj} = \omega_j$, $(j = 1, \dots, n)$, so that (7) tends formally to (6) as $t \rightarrow \infty$, and it is easy to see that (7) is a q -difference equation in the variable $x - b_t$. (Cf. Lemma I, of the Appendix.)

subset \mathcal{I}_0 of \mathcal{U} , and every positive number ϵ , there exists an ordered set of complex numbers z_1, \dots, z_n such that $(\sum_{k=1}^n |z_k - 1|^2)^{1/2} < \epsilon$ and such that for some $Q(\alpha, \mathcal{U})$ solution $y(x, z_1, \dots, z_n)$ of the parametrized equation

$$(8) \quad f(x, z_1 y(x + \omega_1), \dots, z_n y(x + \omega_n)) = 0$$

the inequality $|y(x, z_1, \dots, z_n) - y(x)| < \epsilon$ is valid for every x in \mathcal{I}_0 .

(Obviously every $Q(\alpha, \mathcal{U})$ solution of (6) is a $P(\alpha, \mathcal{U})$ solution of (6). However, the converse is not true. For example, it is almost obvious that the equation $y(x + 1) - y(x) = \phi(x)$ has no $Q(\alpha, \mathcal{U})$ solution whatever if $\phi(x)$ is analytic and different from zero at ∞ , but for such a condition on $\phi(x)$ this equation has, for every α with $|\alpha| < \pi/2$ and every choice of \mathcal{U} as a half-plane $\Re(x) > D > 0$, (with D sufficiently large), a one-parameter family of $P(\alpha, \mathcal{U})$ solutions coinciding with the Nörlund principal solutions.)

PART III. General Program.

The definitions given above are meaningful for every algebraic difference equation (6) with coefficients analytic in a reasonably extensive region. Moreover, special solutions of the approximating q -difference equation

$$(9) \quad f(x, z_1 y(q_{t_1} x + V_1), \dots, z_n y(q_{t_n} x + V_n)) = 0,$$

(used for calculating the primary solutions of (8)), are, for most equations (6) and most choices of z_1, \dots, z_n , readily calculated by recursive relations similar to those appearing in the standard procedure for calculating the Taylor's series coefficients at an ordinary point of an algebraic function. However, the remaining convergence discussions, namely the establishment of a limit as t becomes infinite, and after that the establishment of a limit as the z 's approach unity, has been carried through in this paper only for certain classes of difference equations, treated in Parts V, VI, and VII. These classes are: (Part V) a broad class of linear and non-linear equations, greatly generalizing part of the author's paper cited above, and including in particular most algebraic difference equations with coefficients analytic at ∞ ; (Part VI) all equations of the types appearing in Nörlund's "Differenzenrechnung," §§ 32-36, constituting the core of the Nörlund theory of the principal solution (in the analytic case); (Part VII) a simple type of linear equation, for which the variation of the principal solution with the direction is studied.

Part IV of this paper is devoted to a few simple general theorems, used in the sequel, on the relative invariance of the $P(\alpha, \mathcal{J})$ solution under linear transformation of the independent variable, and under translation of the span.

Part VIII is an appendix in which are collected the definition of a special class of functions, "almost constant functions," useful in Part V, and certain lemmas which are used in Parts V, VI, and VII.

PART IV. On the Relative Invariance of the $P(\alpha, \mathcal{J})$ Solutions, under Linear Transformations of the Independent Variable, and under Translations of the Span.

THEOREM 1. *Given the difference equation*

$$(10) \quad f(x, y(x + \omega_1), \dots, y(x + \omega_n)) = 0.$$

Let $s = Ax + B$, where A and B are any complex numbers, with $A \neq 0$. Let $h(s) = y(x)$, let $v_k = A\omega_k$, ($k = 1, \dots, n$), and let

$$(11) \quad g(s, h(s, v_1), \dots, h(s + v_n)) = 0$$

be the difference equation in $h(s)$ corresponding to (10). (That is,

$$g(s, h_1, \dots, h_n) \equiv f((s - B)/A, h_1, \dots, h_n).$$

Then if $y_0(x)$ is a $P(\alpha, \mathcal{J})$ solution of (10), the function $h_0(s)$ defined by $h_0(s) \equiv y_0((s - B)/A)$ is a $P(\beta, \mathcal{U})$ solution of (11), where $\beta \equiv \alpha + \text{Arg } A \pmod{2\pi}$, and \mathcal{U} is the set of points s such that $(s - B)/A$ is in \mathcal{J} .

Proof. Let (q_t, b_t) , ($t = 1, 2, \dots$), be a $Q(\alpha)$ -sequence, and let $y(x, t)$ be a special solution of

$$(12) \quad f(x, z_1 y(q^{\omega_1} x + V_1), \dots, z_n y(q^{\omega_n} x + V_n)) = 0,$$

where $V_k = b(1 - q^{\omega_k})$, ($k = 1, \dots, n$).¹¹

Let $h(s, t) = y((s - B)/A, t)$. Then $h(s, t)$ is a special solution of

$$(13) \quad g(s, z_1 h(r^{v_1} s + W_1), \dots, z_n h(r^{v_n} s + W_n)) = 0,$$

where $r = q^{(1/A)} = e^{(1/A)\text{Log } q}$, and $W_k = (Ab + B)(1 - r^{v_k})$, ($k = 1, \dots, n$). Since $(r_t, Ab_t + B)$, ($t = 1, 2, \dots$), is a $Q(\beta)$ -sequence (cf. Lemma VI of the Appendix), it follows easily that $y_0((s - B)/A)$ is a $P(\beta, \mathcal{U})$ solution of (11).

¹¹ Whenever it is convenient we omit the subscript t on q_t , b_t , V_{kt} , etc.

THEOREM 2. *Given the difference equation (10). Let c be any complex number. Let $H(x) \equiv y(x - c)$. Let (10) be written in the form*

$$(14) \quad f(x, H(x + c + \omega_1), \dots, H(x + c + \omega_n)) = 0,$$

which will be considered as a difference equation in the unknown function $H(x)$, the spans being $c + \omega_1, \dots, c + \omega_n$. Let $y_0(x)$ be a $P(\alpha, \mathcal{J})$ solution of (10). Then $H_0(x) \equiv y_0(x - c)$ is a $P(\alpha, \mathcal{U})$ solution of (14), where \mathcal{U} consists of the points x such that $x - c$ is in \mathcal{J} .

Proof. Let (q_t, b_t) be a $Q(\alpha)$ -sequence, and let $y(x, t)$ be a special solution of (12). Let $H(x, t) \equiv y(X_1(x, t), t)$, where

$$(15) \quad X_1(x, t) = q^{-c}x + b(1 - q^{-c}).$$

Then $H(x, t)$ is a special solution of

$$(16) \quad f(x, z_1 H(q^{(\omega_1+c)}x + W_1), \dots, z_n H(q^{(\omega_n+c)}x + W_n)) = 0,$$

where $W_k = b(1 - q^{(\omega_k+c)})$, ($k = 1, \dots, n$). Evidently, $\lim_{t \rightarrow \infty} H(x, t) = \lim_{t \rightarrow \infty} y(x - c, t)$, and from this the theorem follows at once.

PART V. A Class of Linear and Non-Linear Equations.

THEOREM 3. *Given the difference equation (10), with $f(x, y_1, \dots, y_n)$ a polynomial of degree d in the indeterminates y_k , the coefficients being functions of x analytic at ∞ ,¹² and the ω 's being complex numbers, with $\omega_1 = 0$, and $\Re(\omega_k) > 0$ when $k > 1$. Let the polynomial in the y_k obtained from $f(x, y_1, \dots, y_n)$ by substituting for each coefficient its limit at ∞ be denoted by $F(y_1, \dots, y_n)$. We shall assume that the following algebraic equation in one unknown γ :*

$$(17) \quad F(\gamma, \dots, \gamma) = 0$$

has exactly d distinct finite roots $\gamma_1, \dots, \gamma_d$, and that for each root γ_j , ($j = 1, \dots, d$), the "limiting separant function"

$$(18) \quad \sum_{k=1}^n F_{y_k}(\gamma_j, \dots, \gamma_j) \sigma^{\omega_k}$$

is different from zero for every σ in the closed interval $[0, 1]$.¹³ (By σ^{ω_k} we understand $e^{\omega_k \text{Log } \sigma}$ if $\sigma \neq 0$, 0 if $\sigma = 0$ and $\omega_k \neq 0$, 1 if $\sigma = 0$ and $\omega_k = 0$.)

¹² The condition of analyticity at ∞ is relaxed in the next theorem.

¹³ Most equations with coefficients analytic at ∞ fulfill these conditions: in fact, if $f(x, y_1, \dots, y_n)$ is any polynomial in the y_k of degree d ; with d distinct finite roots

For every positive D let $\mathcal{H}(D)$ be the half-plane $\Re(x) > D$. Then for every sufficiently large positive D , the $P(0, \mathcal{H}(D))$ solutions of (10) consist of exactly d distinct functions $y_j(x)$, ($j = 1, \dots, d$), each analytic and bounded in $\mathcal{H}(D)$; also $\lim_{\Re(x) \rightarrow +\infty} y_j(x) = \gamma_j$, ($j = 1, \dots, d$); finally, every solution of (10) which is analytic and bounded in a right half-plane $\mathcal{H}(D_1)$ is for some positive D_2 coincident in $\mathcal{H}(D_2)$ with a $P(0, \mathcal{H}(D_2))$ solution.

Proof. We first seek the $Q(0, \mathcal{H}(D))$ solutions of the parametrized equation (8). We shall denote an ordered n -tuple (z_1, \dots, z_n) by the symbol Z , the ordered n -tuple $(1, \dots, 1)$ by the symbol Z_0 . By the distance between $Z = (z_1, \dots, z_n)$ and $Z' = (z'_1, \dots, z'_n)$, written $\delta(Z, Z')$, will be meant $(\sum_{k=1}^n |z_k - z'_k|^2)^{1/2}$. We shall say that the sequence Z', Z'', \dots approaches Z as a limit if $\lim_{n \rightarrow \infty} \delta(Z^{(n)}, Z) = 0$.

It follows from a straightforward continuity argument that there exist positive numbers ϵ, L , such that if $\delta(Z, Z_0) < \epsilon$, then (a) the equation (19) $F(z_1 C, \dots, z_n C) = 0$ has exactly d distinct finite roots C_1, \dots, C_d , and (b) for every $Q(0)$ -sequence $S: (q_t, b_t)$, ($t = 1, 2, \dots$), there is a positive T (depending upon Z and S), such that when $t > T$ the equation

$$(20) \quad f(b_t, z_1 c_0, \dots, z_n c_0) = 0$$

has exactly d distinct finite roots c_0 , and for each such c_0

$$(21) \quad \left| \sum_{k=1}^n z_k f_{y_k}(b, z_1 c_0, \dots, z_n c_0) \dot{q}_t^{\lambda \omega_k} \right| > L$$

for every $\lambda \geq 0$. We shall denote the set of ordered n -tuples Z such that $\delta(Z, Z_0) < \epsilon$ by \mathcal{N} .

We consider (8) for Z in \mathcal{N} . Let $S: (q_t, b_t)$, ($t = 1, 2, \dots$) be a $Q(0)$ -sequence, and let (12) be the corresponding approximating q -difference equation. Let

$$(22) \quad y(x) = \sum_{\lambda} c_{\lambda} u^{\lambda}, \text{ where } u = x - b,$$

$\gamma_1, \dots, \gamma_d$ for the corresponding $F(\gamma, \dots, \gamma)$, and with $F_{y_j}(\gamma_j, \dots, \gamma_j)$ and $F_{y_1}(\gamma, \dots, \gamma_j)$ both different from zero, ($j = 1, \dots, d$), and if $w_1 = 0$ while w_2, \dots, w_{n-1} are chosen in any fashion as points in the right half-plane, then it is evident that there are d analytic curves in the right half-plane, whose equations can easily be written out explicitly, such that if w_n is chosen as any point of the right half-plane not on any one of these curves, then the corresponding difference equation (10) will satisfy all the hypotheses of the theorem.

be a generalized power-series, if there is one, satisfying (12). Let

$$(23) \quad G(u) \equiv f(u + b, z_1 \sum_{\lambda} c_{\lambda} u^{\lambda} q^{\omega_1 \lambda}, \dots, z_n \sum_{\lambda} c_{\lambda} u^{\lambda} q^{\omega_n \lambda}).$$

Then $G(u) \equiv 0$. Evidently c_0 is determined by equating to zero the coefficient of u^0 in $G(u)$; it is any root of equation (20), which for t sufficiently large has d distinct finite roots. We assert that if t is sufficiently large, no non-negative real λ exists for which $c_{\lambda} \neq 0$ and $\lambda \not\equiv 0 \pmod{1}$. For, assuming the contrary, let I be an infinite subsequence of the positive integers, such that for each t in I there exists a non-negative real λ such that $c_{\lambda} \neq 0$ and $\lambda \not\equiv 0 \pmod{1}$, and let $\lambda_0 = \lambda_0(t)$ be for each t in I the smallest such λ .

Then the coefficient of u^{λ_0} in $G(u)$ is $c_{\lambda_0} \sum_{k=1}^n z_k f_{y_k}(b, z_1 c_0, \dots, z_n c_0) q^{\omega_k \lambda_0}$, and this must be zero for each t in I , in contradiction with the properties of \mathcal{N} . Thus

$$(24) \quad y(x) = \sum_{\lambda=0}^{\infty} c_{\lambda} x^{\lambda}, \text{ an ordinary power-series,}$$

if there is any generalized power-series at all which satisfies (12).

For further study of the c_{λ} , ($\lambda = 1, 2, \dots$), we write $f(x, y_1, \dots, y_n)$ in the form

$$(25) \quad f(x, y_1, \dots, y_n) \equiv \sum_{p=1}^m a_p(x) \prod_{s=1}^{s(p)} y_{p,s} - \phi(x),$$

where (for each pair p, s) $y_{p,s}$ is one of the y_k , ($k = 1, \dots, n$), and $1 \leq s(p) \leq d$, and $a_p(x)$ and $\phi(x)$ are analytic at ∞ . Then

$$(26) \quad f(x, z_1 y(q^{\omega_1} x + V_1), \dots, z_n y(q^{\omega_n} x + V_n)) \\ \equiv \sum_{p=1}^m a_p(x) \prod_{s=1}^{s(p)} z_{p,s} y(q_{p,s} x + V_{p,s}) - \phi(x),$$

where $z_{p,s} = z_k$, $\omega_{p,s} = \omega_k$, and $V_{p,s} = V_k$, if $y_{p,s} = y_k$, and where $q_{p,s} = q^{\omega_{p,s}}$. Hence

$$(27) \quad f(x, z_1 y(q^{\omega_1} x + V_1), \dots, z_n y(q^{\omega_n} x + V_n)) \\ \equiv \sum_{p=1}^m a_p(x, Z) \prod_{s=1}^{s(p)} y(q_{p,s} x + V_{p,s}) - \phi(x),$$

where $a_p(x, Z) = a_p(x) \prod_{s=1}^{s(p)} z_{p,s}$.

If t is sufficiently large, then (21) holds, and in the new notation this becomes

$$(28) \quad \left| \sum_{p=1}^m a_p(b, Z) c_0^{s(p)-1} (q^{\lambda}_{p,1} + q^{\lambda}_{p,2} + \cdots + q^{\lambda}_{p,s(p)}) \right| > L,$$

valid for all non-negative real λ .

Let

$$(29) \quad a_p(x, Z) = \sum_{\lambda=0}^{\infty} a_{p,\lambda} u^{\lambda}, \quad (p=1, \cdots, m),$$

and let

$$(30) \quad \phi(x) = \sum_{\lambda=0}^{\infty} \phi_{\lambda} u^{\lambda}.$$

Then

$$(31) \quad \sum_{p=1}^m \left(\sum_{\lambda=0}^{\infty} a_{p,\lambda} u^{\lambda} \right) \prod_{s=1}^{s(p)} \left(\sum_{\lambda=0}^{\infty} c_{\lambda} q^{\lambda}_{p,s} u^{\lambda} \right) = \sum_{\lambda=0}^{\infty} \phi_{\lambda} u^{\lambda},$$

from which the c_{λ} , ($\lambda=1, 2, \cdots$), are determined by equations of the following form

$$(32) \quad c_{\lambda} \sum_{p=1}^m a_{p,0} c_0^{s(p)-1} (q^{\lambda}_{p,1} + \cdots + q^{\lambda}_{p,s(p)}) \\ = \phi_{\lambda} - H_{\lambda}(c_0, c_1, \cdots, c_{\lambda-1}, q_{p,s}, a_{i\theta}),$$

where H_{λ} , ($\lambda=1, 2, \cdots$) is a polynomial, with positive integers for coefficients, in the indicated arguments, (with $i=1, \cdots, m$; $p=1, \cdots, m$; $s=1, \cdots, s(p)$; $\theta=1, \cdots, \lambda$). By virtue of inequality (28), it is evident that if numbers C_{λ} , ($\lambda=1, 2, \cdots$), are defined recursively by the equations

$$(33) \quad LC_{\lambda} = |\phi_{\lambda}| + H_{\lambda}(|c_0|, C_1, \cdots, C_{\lambda-1}, 1, |a_{i\theta}|),$$

then (34) $|c_{\lambda}| \leq C_{\lambda}$, ($\lambda=1, 2, \cdots$). (We use here the fact, which follows immediately from Lemma IV, that $|q_{p,s}| < 1$.) If we define $C(u)$ to be $\sum_{\lambda=1}^{\infty} C_{\lambda} u^{\lambda}$, then (33) are the determining relations for the coefficients of an analytic function $C(u)$ which vanishes at $u=0$ and satisfies the equation

$$(34) \quad LC(u) = \sum_{\lambda=1}^{\infty} |\phi_{\lambda}| u^{\lambda} + \sum_{p=1}^m \left\{ \left[\sum_{\lambda=1}^{\infty} |a_{p,\lambda}| u^{\lambda} \right] [|c_0| + C(u)]^{s(p)} \right. \\ \left. + |a_{p,0}| [(|c_0| + C(u))^{s(p)} - |c_0|^{s(p)} - s(p) |c_0|^{s(p)-1} C(u)] \right\}.$$

Let ϵ_0 be any positive number. Then there exists a positive number D independent of t such that if t is sufficiently large, and $|u| < |b| - D$, then $\left| \sum_{\lambda=1}^{\infty} |\phi_{\lambda}| u^{\lambda} \right| < \epsilon_0$, and $\left| \sum_{\lambda=1}^{\infty} |a_{p,\lambda}| u^{\lambda} \right| < \epsilon_0$, ($p=1, \cdots, m$). (See Lemmas XVII and XIV, and Definitions I, II, of the Appendix.) Thus if D is a

sufficiently large positive number, (independent of t), then for each sufficiently large t there is a j in the set $(1, \dots, d)$, such that the coefficients in (34), considered as an equation in $C(u)$, are arbitrarily near the coefficients in the following equation in X :

$$(35) \quad LX = \sum_{p=1}^m A_p(Z) \{ [|\gamma_j(Z)| + X]^{s(p)} - |\gamma_j(Z)|^{s(p)} - s(p) |\gamma_j(Z)|^{s(p)-1} X \},$$

where $A_p(Z) = \lim_{x \rightarrow \infty} |a_p(x, Z)|$, and $\gamma_j(Z)$, ($j = 1, \dots, d$), are the d roots C of (19). Since $C(u)$ vanishes at $u = 0$, it follows that, throughout the region $|u| < |b| - D$, $C(u)$ is, for D sufficiently large, arbitrarily near the solution $X = 0$ of (35). Since the equation in X obtained from (35) by differentiation with respect to X is not satisfied at $X = 0$, $C(u)$ is analytic throughout the region $|u| < |b| - D$, if D is sufficiently large.

Hence it follows from (34) and the fact that c_0 is near one or other of the $\gamma_j(Z)$ when t is large, that there exist positive numbers D , M , independent of t and Z , such that whenever Z is in \mathcal{N} , and S is any $Q(0)$ -sequence, then when t is sufficiently large the special solutions $\sum_{\lambda=0}^{\infty} c_{\lambda} u^{\lambda}$ of the corresponding approximating q -difference equation (12) satisfy the inequality

$$(36) \quad \left| \sum_{\lambda=0}^{\infty} c_{\lambda} u^{\lambda} \right| < M$$

for all u such that $|u| < |b| - D$. Hence if $y(x, Z)$ is any $Q(0, \mathcal{H}(D_1))$ solution of (8), and if $D_2 = \max(D, D_1)$, then

$$(37) \quad |y(x, Z)| \leq M$$

throughout the half-plane $\mathcal{H}(D_2)$ (Cf. Lemma XVIII.) Hence if $y(x)$ is any $P(0, \mathcal{H}(D_1))$ solution of (10), and if $D_2 = \max(D, D_1)$, then

$$(38) \quad |y(x)| \leq M$$

throughout the half-plane $\mathcal{H}(D_2)$.

Since for every sufficiently large t we have a special solution of (12) satisfying (36) in the region $|u| < |b| - D$, it follows from the standard compactness argument for bounded families of analytic functions that as t becomes infinite on a suitable subsequence I of the positive integers, $\sum_{\lambda=0}^{\infty} c_{\lambda}(x-b)^{\lambda}$ approaches a limit for every x in the half-plane $\mathcal{H}(D)$, uniformly in every closed bounded subset of that half-plane. Such a limit

function $y_0(x, Z)$ is plainly a $Q(0, \mathcal{A}(D))$ solution of (8). It is majorized in $\mathcal{A}(D)$ by the constant M , and moreover, for some j in the set $(1, \dots, d)$ we have

$$(39) \quad \lim_{\Re(x) \rightarrow +\infty} y_0(x, Z) = \gamma_j(Z).$$

For let ϵ_0 be any positive number, and let D_1 be a positive number greater than D and so large that $|C(u)| < \epsilon_0$ if $|u| < |b| - D_1$ and t is sufficiently large. Let x_0 be any point such that $\Re(x_0) > D_1$. Let $u_0 = x_0 - b$.

Let T be so large that $|y_0(x_0, Z) - \sum_{\lambda=0}^{\infty} c_\lambda u_0^\lambda| < \epsilon_0$ if $t > T$, and so large that $|c_0 - \gamma_j(Z)| < \epsilon_0$ if $t > T$ and j is properly chosen and so large that $|C(u)| < \epsilon_0$ if $|u| < |b| - D_1$ and $t > T$. Then

$$\begin{aligned} |y_0(x_0, Z) - \gamma_j(Z)| &\leq |y_0(x_0, Z) - \sum_{\lambda=0}^{\infty} c_\lambda u_0^\lambda| \\ &\quad + \left| \sum_{\lambda=1}^{\infty} c_\lambda u_0^\lambda \right| + |c_0 - \gamma_j(Z)| \leq 3\epsilon_0. \end{aligned}$$

Since c_0 can be chosen, for large t , near any desired one of the $\gamma_j(Z)$ it is easy to see that for any desired j a $y_0(x, Z)$ may be obtained such that $y_0(x, Z)$ tends to $\gamma_j(Z)$ as $\Re(x)$ tends to positive infinity. Let $y_j(x, Z)$, ($j = 1, \dots, d$) be such limit functions $y_0(x, Z)$ with $\lim_{\Re(x) \rightarrow +\infty} y_j(x, Z) = \gamma_j(Z)$, ($j = 1, \dots, d$).

The functions $y_j(x, Z_0)$, ($j = 1, \dots, d$), are $Q(0, \mathcal{A}(D))$, hence $P(0, \mathcal{A}(D))$ solutions of (10), having all the properties of the $y_j(x)$ asserted in the statement of this theorem, except possibly the uniqueness property.

We assert now that there exist d functions $Y_j(x, b)$, ($j = 1, \dots, d$), defined for all sufficiently large positive b , and for all x satisfying the condition $|x - b| < b - D_2$, for some positive D_2 independent of b , such that for every solution $y^*(x)$ of (10), bounded and analytic in a right half-plane $\mathcal{A}(D_3)$, there is a j in the set $(1, \dots, d)$ such that $y^*(x) = \lim_{b \rightarrow \infty} Y_j(x, b)$ in some half-plane $\mathcal{A}(D_4)$, the limit being uniform in every closed bounded subset of $\mathcal{A}(D_4)$. The justification of this assertion would evidently complete the proof of the theorem.

Let $\mathcal{B}_1, \dots, \mathcal{B}_d$ be disjoint neighborhoods of $\gamma_1, \dots, \gamma_d$, respectively. Let b_0 be a positive number such that if $b > b_0$, then there is exactly one root c_0 of

$$(40) \quad f(b, c_0, \dots, c_0) = 0$$

in each \mathcal{B}_j , ($j = 1, \dots, d$), and for each such c_0

$$(41) \quad \left| \sum_{k=1}^n f_{v_k}(b, c_0, \dots, c_0) (1 - b^{-1})^{\lambda \omega_k} \right| > L,$$

for every $\lambda \geq 0$. (The existence of such a b_0 follows by continuity from the hypotheses of this theorem.)

For all b greater than b_0 we define q as the function $1 - b^{-1}$, and we define $Y_j(x, b)$, as that special solution of

$$(42) \quad f(x, y(q^{\omega_1}x + V_1), \dots, y(q^{\omega_n}x + V_n)) = 0$$

for which $Y_j(b, b)$ lies in \mathcal{B}_j , ($j = 1, \dots, d$). By the discussion given above, $Y_j(x, b)$ is almost constant in direction 0. (Cf. Appendix, Def. II.)

Let $y^*(x)$ be any solution of (10), bounded and analytic in a right half-plane $\mathcal{H}(D_s)$. Let b be a positive number greater than D_s , and greater than b_0 . Then $y^*(x)$ is a solution of the q -difference equation

$$(43) \quad f(x, y(q^{\omega_1}x + V_1), \dots, y(q^{\omega_n}x + V_n)) + E(x, b) = 0,$$

where

$$(44) \quad E(x, b) = f(x, y^*(x + \omega_1), \dots, y^*(x + \omega_n)) \\ - f(x, y^*(q^{\omega_1}x + V_1), \dots, y^*(q^{\omega_n}x + V_n)).$$

Let $y(x) = \sum_{\lambda=0}^{\infty} y_{\lambda} u^{\lambda}$, where $u = x - b$. Then by (43) we have

$$(45) \quad f(b, y_0, \dots, y_0) + E(b, b) = 0.$$

Now $E(b, b) = f(b, y^*(b + \omega_1), \dots, y^*(b + \omega_n)) - f(b, y^*(b), \dots, y^*(b))$. Since $y^*(x)$ is bounded in a right half-plane, it follows from Lemma XX, (with $x = b$), that $y^*(b + \omega_k) - y^*(b)$ is small when b is large, ($k = 1, \dots, n$). Hence $E(b, b)$ is small when b is large. From this it follows that if b is large then y_0 is in the union $\mathcal{B}_1 + \dots, \mathcal{B}_d$, and since y_0 varies continuously with b , there is a fixed J in $(1, \dots, d)$ such that y_0 is in \mathcal{B}_J for all large b .

Let $Y_J(x, b) = \sum_{\lambda=0}^{\infty} Y_{\lambda} u^{\lambda}$. Let $s(x) = y^*(x) - Y_J(x, b) = \sum_{\lambda=0}^{\infty} s_{\lambda} u^{\lambda}$. Let $x_k = q^{\omega_k}x + V_k$, ($k = 1, \dots, n$). Then $f(x, s(x_1) + Y_J(x_1, b), \dots, s(x_n) + Y_J(x_n, b)) + E(x, b) = 0$. Since $f(x, Y_J(x_1, b), \dots, Y_J(x_n, b)) = 0$, we conclude that $s(x)$ is a solution of the q -difference equation

$$(46) \quad G(x, s(x_1), \dots, s(x_n)) = 0,$$

where

$$G(x, s_1, \dots, s_n) \equiv f(x, s_1 + Y_J(x_1, b), \dots, s_n + Y_J(x_n, b)) \\ - f(x, Y_J(x_1, b), Y_J(x_n, b)) + E(x, b).$$

Thus $G(x, s_1, \dots, s_n)$ is a polynomial in the s_k , with coefficients functions of x and b which are almost constant in direction 0. (That $E(x, b)$ is almost constant in direction 0 is proved in Lemma XXII. The almost constancy of the other coefficients follows from Lemma XV.)

We note that s_0 is small when t is large.

Let $\sigma_k = \sum_{\lambda=1}^{\infty} s_{\lambda} q^{\lambda \omega_k} u^{\lambda}$, ($k = 1, \dots, n$). Then

$$(47) \quad G(x, s(x_1), \dots, s(x_n)) = \sum_{k=1}^n f_{y_k}(b, s_0 + Y_0, \dots, s_0 + Y_0) \sigma_k \\ + \sum_{k=1}^n [f_{y_k}(x, s_0 + Y_J(x_1, b), \dots, s_0 + Y_J(x_n, b)) \\ - f_{y_k}(b, s_0 + Y_0, \dots, s_0 + Y_0)] \sigma_k \\ + \sum' H(i_1, \dots, i_n; x; b) \sigma_1^{i_1} \dots \sigma_n^{i_n} + K(x, b),$$

where \sum' is a summation over all positive integers i_1, \dots, i_n such that $2 \leq i_1 + \dots + i_n \leq d$, where the $H(i_1, \dots, i_n; x; b)$ and $K(x, b)$ are almost constant in direction 0, and moreover for some fixed D_4 , $K(x, b)$ is arbitrarily small if b is sufficiently large, throughout the region $|x - b| < b - D_4$. (Cf. Lemma XXII, and note that

$$K(x, b) = E(x, b) + f(x, s_0 + Y_J(x_1), \dots, s_0 + Y_J(x_n)) \\ - f(x, Y_J(x_1), \dots, Y_J(x_n)).$$

Thus, if

$$W_k(x, b) = f_{y_k}(x, s_0 + Y_J(x_1, b), \dots, s_0 + Y_J(x_n, b)), \quad (k = 1, \dots, n),$$

we have

$$(48) \quad \sum_{k=1}^n f_{y_k}(b, s_0 + Y_0, \dots, s_0 + Y_0) \sum_{\lambda=1}^{\infty} s_{\lambda} u^{\lambda} q^{\lambda \omega_k} = - \sum_{k=1}^n (W_k(x, b) \\ - W_k(b, b)) \sigma_k - \sum' H(i_1, \dots, i_n; x; b) \sigma_1^{i_1} \dots \sigma_n^{i_n} - K(x, b).$$

We observe that $W_k(x, b)$ is almost constant in direction 0. Let

$$W_k(x, b) = \sum_{\lambda=0}^{\infty} W_{k,\lambda} u^{\lambda}, \quad H(i_1, \dots, i_n; x; b) = \sum_{\lambda=0}^{\infty} H_{\lambda}(i_1, \dots, i_n) u^{\lambda}, \\ K(x, b) = \sum_{\lambda=0}^{\infty} K_{\lambda} u^{\lambda}.$$

Then the s_{λ} are determined, ($\lambda = 1, 2, \dots$), by equations of the form

$$(49) \quad s_\lambda \sum_{k=1}^n f_{y_k}(b, s_0 + Y_0, \dots, s_0 + Y_0) q^{\lambda \omega_k} \\ = -N_\lambda(s \theta q^{\theta \omega_k}; W_{k,r}; H_p(i_1, \dots, i_n); K_h),$$

($r, \theta = 1, \dots, \lambda - 1$; $k = 1, \dots, n$; $p = 1, \dots, \lambda - 2$; $h = 1, \dots, \lambda$), where N_λ is a polynomial in the indicated arguments, with positive coefficients.

It follows from a straightforward continuity argument that if b is sufficiently large

$$| \sum_{k=1}^n f_{y_k}(b, s_0 + Y_0, \dots, s_0 + Y_0) q^{\lambda \omega_k} | > L' = L/2$$

for all positive λ . Thus $|s_\lambda| \leq S_\lambda$, where the numbers S_λ , ($\lambda = 1, 2, \dots$) are defined recursively by

$$(50) \quad L'S_\lambda = N_\lambda(S_\theta; |W_{k,r}|; |H_p(i_1, \dots, i_n)|; |K_h|).$$

Evidently if $S(u) = \sum_{\lambda=1}^{\infty} S_\lambda u^\lambda$, then $S(u)$ satisfies the algebraic equation

$$(51) \quad L'S(u) = \sum_{k=1}^n (W_k^A(x, b) - W_k^A(b, b)) S(u) \\ + \Sigma' H^A(i_1, \dots, i_n; x; b) S^{(i_1 + \dots + i_n)}(u) + K^A(x, b),$$

where the superscript A indicates here the operation defined by the equation

$$(\sum_{\lambda=0}^{\infty} F_\lambda u^\lambda)^A = \sum_{\lambda=0}^{\infty} |F_\lambda| u^\lambda. \text{ Let limit } H^A(i_1, \dots, i_n; b; b) = h(i_1, \dots, i_n).$$

If D_4 is large, the coefficients in (51), considered as an equation in $S(u)$, are for all large b and all x satisfying $|x - b| < b - D_4$, arbitrarily near the coefficients of the following equation in X :

$$(52) \quad L'X = \Sigma' h(i_1, \dots, i_n) X^{(i_1 + \dots + i_n)}.$$

Hence, since the equation obtained from (52) by differentiation with respect to X is not satisfied when $X = 0$, it follows that if D_4 is fixed as a sufficiently large positive number, then $S(u)$ as defined by (75) and the condition $S(0) = 0$ is analytic in $|u| < b - D_4$, and sufficiently small so that

$$(53) \quad | \Sigma' H^A(i_1, \dots, i_n; x; b) S^{(i_1 + \dots + i_n - 1)}(u) | < L'/4,$$

and

$$(54) \quad | \Sigma (W_k^A(x, b) - W_k^A(b, b)) | < L'/4, \text{ in } |u| < b - D.$$

Let D_4 be fixed to satisfy these conditions. From (51) we have

$$(55) \quad S(u) = K^A(x, b) / [L' - \Sigma(W_k^A(x) - W_k^A(b)) \\ - \Sigma'H^A(i_1, \dots, i_n; x; b) S^{(i_1 + \dots + i_n - 1)}(u)],$$

so that $|S(u)| < 2K^A(x, b)/L'$ in $|u| < b - D_4$, and since $K^A(x, b)$ is arbitrarily small in $|x - b| < b - D_4$, if b is large, and since s_0 is small when b is large, $|s(x)|$ is small throughout $|x - b| < b - D_4$ if b is large. That is, $y(x) = \lim_{b \rightarrow \infty} Y_J(x, b)$, the limit being uniform in every closed bounded subset of the half-plane $\mathcal{H}(D_4)$. As observed above, this completes the proof of the theorem.

THEOREM 4. *If in the hypotheses of Theorem 3 the phrase "analytic at ∞ " is replaced by "almost constant in direction 0," and "limit at ∞ " is replaced by "limit as $\Re(x)$ becomes positively infinite," the conclusions remain valid without change, and in addition, for sufficiently large positive D , each $P(0, \mathcal{H}(D))$ solution is almost constant in direction 0. (We emphasize the fact, which was used in the proof of Theorem 3, and is proved in Lemma XVII, that every function which is analytic at ∞ is almost constant in direction 0. The $P(0, \mathcal{H}(D))$ solutions of equation (10) obtained in this paper constitute a large class of functions which are almost constant in direction 0 and which are usually not analytic at ∞ .)*

Proof. The proof is essentially identical with the proof of Theorem 3, since the salient property of functions analytic at ∞ which was used in the proof of Theorem 3 is the property of being almost constant in direction 0.

PART VI. The Nörlund Equations.

The equations to be considered are (1) and (3), under the following assumptions:

Either (56) $\phi(x)$ is an entire function, and for some positive numbers C, K the inequality $|\phi(x)| < Ce^{K|\omega|}$ is valid for all x , K being less than $2\pi/|\omega|$ in the case of equation (1), less than $\pi/|\omega|$ in the case of equation (3);

Or (57) $\phi(x)$ is analytic at every point of a sector \mathcal{S} defined by inequalities $(-\pi/2) \leq M < \arg x < L \leq (\pi/2)$, and analytic on the boundary of \mathcal{S} . Also, ω lies in \mathcal{S} . Finally, if $\beta_0 = \min(\pi/2, L - \arg \omega, \arg \omega - M)$, then there exist positive numbers C, K such that the inequality $|\phi(x)| < Ce^{K|\omega|}$ is valid for all x in \mathcal{S} , K being less than $(2\pi/|\omega|) \sin \beta_0$.

in the case of equation (1), less than $(\pi/|\omega|) \sin \beta_0$ in the case of equation (5).

Under these assumptions the Nörlund principal solution exists in \mathcal{D} and has a representation (for equation (1))

$$(58) \quad f(x) = (\pi/2i) \int_{C_0} \psi(x + \omega z) \csc^2 \pi z \, dz,$$

and has a representation (for equation (3)),

$$(59) \quad f(x) = -i \int_{C_0} \phi(x + \omega z) \csc \pi z \, dz,$$

where $\psi(x) = \int_{\alpha}^x \phi(z) \, dz$, (α being an arbitrary number in the open interval $(-1, 0)$, and the path of integration from α to x being a straight line segment), and where C_0 is the contour, (described in the sense of increasing ρ), $z = \alpha + \rho e^{i\beta_0}$ ($-\infty < \rho \leq 0$), $z = \alpha + \rho e^{-i\beta_0}$ ($0 \leq \rho < \infty$). We define $\beta_0 = \pi/2$ in the case where $\phi(x)$ is entire.)¹⁴

In what follows we shall prove

THEOREM 5. *If either (56) or (57) is satisfied, then the $P(\text{Arg } \omega, \mathcal{D})$ solutions of (1) and (3) exist and coincide with the Nörlund principal solutions.*

Proof. Section A. The $P(\text{Arg } \omega, \mathcal{D})$ solutions of equation (1). We change the notation slightly, writing (1) in the form

$$(60) \quad y(x + \omega) - y(x) = \omega \phi(x).$$

Let

$$(61) \quad z_1 y(x + \omega) - z_0 y(x) = \omega \phi(x)$$

be the parametrized equation corresponding to (60). Let

$$(62) \quad (q_t, b_t), \quad (t = 1, 2, \dots),$$

be a $Q(\text{Arg } \omega)$ -sequence. Let

$$(63) \quad z_1 y(q^\omega x + V) - z_0 y(x) = \omega \phi(x)$$

be an approximating q -difference equation for (61) and (62). (We omit the subscript t ; $V = b(1 - q^\omega)$.) Let

¹⁴ We have paraphrased Nörlund's statements, for the sake of brevity; the essential content of Nörlund's original statements, as well as of the generalizations which he suggests, is contained in this modified version.

$$(64) \quad y(x) = \sum_{\lambda} c_{\lambda} u^{\lambda}, \quad (\text{where } u = x - b),$$

be a generalized power-series, if there is one, satisfying (63) in a quasi-neighborhood of $x = b$. Let

$$(65) \quad \phi(x) = \sum_{k=0}^{\infty} \phi_k u^k.$$

Then

$$(66) \quad z_1 \sum_{\lambda} c_{\lambda} q^{\omega \lambda} u^{\lambda} - z_0 \sum_{\lambda} c_{\lambda} u^{\lambda} = \omega \sum_{k=0}^{\infty} \phi_k u^k.$$

Hence

$$(67) \quad c_{\lambda} (z_1 q^{\omega \lambda} - z_0) = \omega \phi_{\lambda}$$

for every λ , where we define ϕ_{λ} to be zero if λ is not a non-negative integer.

We consider next the equation

$$(68) \quad z_1 q^{\omega \lambda} - z_0 = 0.$$

By Lemma V the existence for each sufficiently large t of a non-negative λ satisfying (68) necessitates that either $z_0 = z_1$, or that

$$(69) \quad [\text{Arg } (z_0/z_1) + 2\pi k_0] / [|\text{Log } (z_0/z_1) + 2\pi k_0 i|] = 0,$$

and

$$(70) \quad [\text{Log } |z_0/z_1|] / [|\text{Log } (z_0/z_1) + 2\pi k_0 i|] = -1$$

for some integer k_0 . Equations (69) and (70) imply, for z_0 and z_1 near 1, that $k_0 = 0$. Thus $\text{Arg } (z_0/z_1) = 0$, and $\text{Log } |z_0/z_1| = -|\text{Log } (z_0/z_1)|$, so that $0 < (z_0/z_1) < 1$. By Lemma V again, $\text{Arg Log } (q^{\omega}) = \pi$, whence $0 < q^{\omega} < 1$.

We distinguish several cases, according to whether equation (68) has for every sufficiently large t a non-negative solution in λ . If $z_0 = z_1$, equations (67) have no solution unless $\phi(b) = 0$. We shall put this case aside temporarily.

First Case. $0 < z_0/z_1 < 1$, and, for every sufficiently large t , $0 < q^{\omega} < 1$. Then for each sufficiently large t there is exactly one non-negative λ_0 such that $z_1 q^{\omega \lambda_0} - z_0 = 0$, namely $\lambda_0 = [\text{Log } (z_0/z_1)] / [\text{Log } (q^{\omega})]$. Thus if equations (67) have any solution at all in the c_{λ} , then in that solution c_{λ_0} is arbitrary. (If $\phi_{\lambda_0} \neq 0$, there is no solution; if $\phi_{\lambda_0} = 0$, then c_{λ_0} is arbitrary.) The case where there is no solution at all can obviously be avoided by suitable choice of the $Q(\text{Arg } \omega)$ -sequence. For example, if R_1, R_2, \dots is a sequence of numbers such that $0 < R_t < 1$, and limit $R_t = 1$, with R_t

distinct from every k -th root of z_0/z_1 , ($k = 1, 2, \dots$), and if $q_t = R_t^{(1/\omega)} = e^{(1/\omega) \log R_t}$, then for some sequence b_1, b_2, \dots the sequence of pairs (q_t, b_t) , ($t = 1, 2, \dots$), is a $Q(\text{Arg } \omega)$ -sequence (by Lemma VII), and for this choice of (62) the equations (67) evidently do have a solution. We confine our attention to the case where there is a solution in the c_λ for equations (67) (of course the other case makes no contribution to the principal solution). Then $c_k = (\omega \phi_k) / (z_1 q^{\omega k} - z_0)$, (all non-negative k such that $z_1 q^{\omega k} - z_0 \neq 0$), and $c_\lambda =$ an arbitrary constant when $\lambda = \lambda_0$. Hence

$$(71) \quad y(x) = \omega \Sigma' \phi_k u^k (z_1 q^{\omega k} - z_0)^{-1} + c_{\lambda_0} u^{\lambda_0},$$

(where Σ' indicates the sum for all non-negative integers different from λ_0 , which may conceivably be an integer). We shall now transform (71) into an expression for $y(x)$ as a contour integral, following steps analogous to those used by Nörlund. Some of the steps will not be justified unless the impossible restriction $|z_1/z_0| < 1$ is made. However, these steps will be carried through formally, as a heuristic device, and the final expression for $y(x)$ will be shown to be valid.

Let $\beta_0 = \pi/2$ in the case (56). Let $\beta_0 = \min(\pi/2, L - \text{Arg } \omega, \text{Arg } \omega - M)$ in the case (57). In either case let β be such that $\beta < \beta_0$, but at the same time $(2\pi \sin \beta)/|\omega| > K$.

We have, formally, from (71),

$$\begin{aligned} (72) \quad y(x) &= (-\omega/z_0) \Sigma' \phi_k u^k (1 - [z_1 q^{\omega k}/z_0])^{-1} + c_{\lambda_0} u^{\lambda_0} \\ &= (-\omega/z_0) \Sigma' [\phi_k u^k \sum_{n=0}^{\infty} (z_1 q^{\omega k}/z_0)^n] + c_{\lambda_0} u^{\lambda_0} \\ &= (-\omega/z_0) \sum_{n=0}^{\infty} [(z_1/z_0)^n (\Sigma' \phi_k q^{\omega n k} u^k)] + c_{\lambda_0} u^{\lambda_0} \\ &= (-\omega/z_0) \sum_{n=0}^{\infty} [(z_1/z_0)^n \phi(x_n)] + c_{\lambda_0} u^{\lambda_0}, \end{aligned}$$

where $x_0 = x$ and $x_{n+1} = q^\omega x_n + V$, ($n = 0, 1, \dots$). Hence

$$\begin{aligned} (73) \quad y(x) &= (-\omega/z_0) (1/2\pi i) \int_C (z_1/z_0)^\xi \phi(q^{\omega \xi}(x - b) + b) \pi \cot \pi \xi d\xi \\ &\quad + (-\omega/z_0) \phi(x) + c_{\lambda_0} u^{\lambda_0},^{15} \end{aligned}$$

where C is the contour (described in the sense of increasing ρ), $\xi = E - \rho e^{i\beta}$, ($-\infty < \rho \leq 0$), $\xi = E + \rho e^{-i\beta}$, ($0 < \rho < \infty$), with $0 < E < 1$, and where

¹⁵ Cf. Nörlund, *loc. cit.*, p. 69. We have translated the contours one unit to the right to simplify the problem of keeping the variable of integration within the domain of analyticity of $\phi(x)$.

$(z_0/z_1)^{\xi} = e^{\xi \operatorname{Log}(z_0/z_1)}$. This integral may be transformed in turn (using the methods of Nörlund again), to

$$\begin{aligned}
 (74) \quad y(x) = & (\omega/z_0) \int_{C_1} (z_1/z_0)^{\xi} \phi(q^{\omega\xi}(x-b) + b) (1 - e^{-2\pi i\xi})^{-1} d\xi \\
 & + (\omega/z_0) \int_{C_2} (z_1/z_0)^{\xi} \phi(q^{\omega\xi}(x-b) + b) (1 - e^{2\pi i\xi})^{-1} d\xi \\
 & - (\omega/z_0) \int_{C_3} (z_1/z_0)^{\xi} \phi(q^{\omega\xi}(x-b) + b) d\xi \\
 & + (-\omega/z_0) \phi(x) + c_{\lambda_0} u^{\lambda_0},
 \end{aligned}$$

where C_1 is the contour $\xi = E + \rho e^{i\beta}$, ($0 \leq \rho < \infty$), C_2 is the contour $\xi = E + \rho e^{-i\beta}$, ($0 \leq \rho < \infty$), and C_3 is the contour $\xi = E + \rho$, ($0 \leq \rho < \infty$). Employing the change of variables $\theta = q^{\omega\xi}(x-b) + b$ in the integral over C_3 , and using the fact that c_{λ_0} is arbitrary, one obtains

$$\begin{aligned}
 (75) \quad y(x) = & (\omega/z_0) \int_{C_1} (z_1/z_0)^{\xi} \phi(q^{\omega\xi}(x-b) + b) (1 - e^{-2\pi i\xi})^{-1} d\xi \\
 & + (\omega/z_0) \int_{C_2} (z_1/z_0)^{\xi} \phi(q^{\omega\xi}(x-b) + b) (1 - e^{2\pi i\xi})^{-1} d\xi \\
 & + (z_0 \operatorname{Log} q)^{-1} u^{\lambda_0} \int_E^{x'} (\theta - b)^{-\lambda_0-1} \phi(\theta) d\theta \\
 & + (-\omega/z_0) \phi(x) + cu^{\lambda_0},
 \end{aligned}$$

where $x' = q^{\omega E}(x-b) + b$, and c is an arbitrary constant. We shall show now that if z, z_1 are sufficiently near 1, 1, then (75) is meaningful, and obtainable from (71) by analytic continuation, provided x is in \mathcal{D} , and t is sufficiently large. In fact, we shall show that for every closed bounded subset \mathcal{F} of \mathcal{D} , there is a positive T such that if $t > T$ then (75) is meaningful and obtainable from (71) by analytic continuation, for every x in \mathcal{F} .

We consider first whether $q^{\omega\xi}(x-b) + b$ is in the domain of analyticity of $\phi(x)$. It suffices to consider the case (57). Let \mathcal{F} be a closed bounded subset of \mathcal{D} . Since $x' = q^{\omega E}x + b(1 - q^{\omega E})$, it is evident that if t is sufficiently large, then x' is arbitrarily near $x + \omega E$ throughout \mathcal{F} . Since the set $\mathcal{F} + \omega E$ is at a positive distance from the boundary of \mathcal{D} , it follows that if t is sufficiently large then x' is in \mathcal{D} for every x in \mathcal{F} . Now on C_1 and C_2 $q^{\omega\xi}(x-b) + b = q^{\omega\rho(\cos\beta \pm i\sin\beta)}(x'-b) + b$, so since x' is in \mathcal{D} , and $0 < \beta < \beta_0$, we may apply Lemma XIII to show that $q^{\omega\xi}(x-b) + b$ is in \mathcal{D} , if ξ is on C_1 or C_2 .

Now the integrals in (75), for any fixed β such that $(2\pi \sin \beta)/|\omega| > K$,

converge if z_1 and z_2 are sufficiently near unity. (We note that $\phi(q^{\omega\zeta}(x-b) + b)$ is bounded when ζ is on C_1 and C_2 , since $q^{\omega\zeta}(x-b) + b$ tends to b as a limit when ζ becomes infinite on C_1 or C_2 .) We assert that if z_1, z_0 are sufficiently small $y(x)$ as defined by (75) coincides in a quasi-neighborhood of b with $y(x)$ as defined by (71). It suffices to prove that $y(x)$ as defined by (75) is, in a quasi-neighborhood of b , a generalized power-series satisfying (63), since every such function is given by (71). That (75) is a generalized power-series is immediately apparent. That it defines $y(x)$ as a solution of (63) follows from the fact that $z_1 y(q^{\omega}x + V) - z_0 y(x)$, calculated from (75), is easily seen by a familiar argument in the calculus of residues to be equal to $\omega\phi(x)$.

Thus $y(x)$ as defined by (75) is an analytic continuation of $y(x)$ as defined by (71), and it is plain that for every point of \mathcal{F} the analytic continuation is an analytic continuation along a radius $\arg(x-b) = \text{constant}$.¹⁶ We shall use the symbol $y(x; t; z_0, z_1)$ to denote $y(x)$ as defined by (75). Then $y(x; t; z_0, z_1)$ is a special solution of (63), defined in a region \mathcal{B}_t , ($t = 1, 2, \dots$), such that if \mathcal{F} is any closed bounded subset of \mathcal{S} , then for all sufficiently large t \mathcal{F} is included in \mathcal{B}_t .

We assert that if the arbitrary constant c is specialized as a suitable function of t , then $\lim_{t \rightarrow \infty} y(x; t; z_0, z_1)$ exists for every x in \mathcal{S} , uniformly in every closed bounded subset of \mathcal{S} , and that every limit function $y(x; z_0, z_1)$ is given by the equation

$$\begin{aligned} (76) \quad y(x; z_0, z_1) = & (\omega/z_0) \int_{C_1} (z_1/z_0)^{\zeta} \phi(x + \omega\zeta) (1 - e^{-2\pi i \zeta})^{-1} d\zeta \\ & + (\omega/z_0) \int_{C_2} (z_1/z_0)^{\zeta} \phi(x + \omega\zeta) (1 - e^{2\pi i \zeta})^{-1} d\zeta \\ & + (1/z_0) (z_0/z_1)^{(x/\omega)} \int_E^{x''} (z_0/z_1)^{(-\theta/\omega)} \phi(\theta) d\theta \\ & + (-\omega/z_0) \phi(x) + c' (z_0/z_1)^{(x/\omega)}, \end{aligned}$$

where $x'' = x + E\omega$, and c' is arbitrary.

We omit the proof, which is completely straightforward. Estimates of $|q^{\omega\zeta}(x-b) + b|$ and $|(q^{\omega\zeta}(x-b) + b) - (x + \omega\zeta)|$, useful for the details of the verification, are given in Lemmas X and XI. We remark also that one satisfactory method of choosing c is in accordance with the formula $c = c' e^{-\lambda_0 \text{Log}(-b)}$.

Thus $y(x; z_0, z_1)$ is a $Q(\text{Arg } \omega, \mathcal{S})$ solution of (61); conversely every $Q(\text{Arg } \omega, \mathcal{S})$ solution of (61), obtained under the First Case, is given by (76).

¹⁶ Whenever $q^{\omega\zeta}(x-b) + b$ is in \mathcal{S} , so is $q^{\omega\zeta}(x'-b) + b$ for every x' on the line segment joining b to x .

We consider next the *Second Case*, where either z_0/z_1 , or q^ω for infinitely many t , lies outside the open interval $(0, 1)$. Conceivably $Q(\text{Arg } \omega, \mathcal{J})$ solutions may exist, not given by (76). To show that this does not happen, it suffices to observe that in the Second Case the generalized power-series (71), (in this Case possibly deprived of the term $c_{\lambda_0} u^{\lambda_0}$), has an analytic continuation (75) with a suitable, (not necessarily arbitrary), c . (This is verified as in the First Case.) Hence if there is a limit for $y(x)$, (as given by (75)), as t becomes infinite, (usually there is no limit, but because of the results of the First Case this is a matter of indifference), then it is given by (76).

Next, let $R: (z'_0, z'_1), (z''_0, z''_1), \dots$ be a sequence of pairs (z_0, z_1) such that $z_0^{(n)}$ and $z_1^{(n)}$ tend to unity as n becomes infinite. We assert that if R is suitably chosen, (namely, to satisfy the condition $0 < (z_0^{(n)}/z_1^{(n)}) < 1$ for every n), then c' can be chosen as a function of n such that $y(x; z_0^{(n)}, z_1^{(n)})$ tends to a limit function $F(x; R)$, uniformly in every closed bounded subset of \mathcal{J} , the limit function being given by

$$\begin{aligned}
 (77) \quad F(x; R) = & \omega \int_{C_1} \phi(x + \omega\xi) (1 - e^{-2\pi i\xi})^{-1} d\xi \\
 & + \omega \int_{C_2} \phi(x + \omega\xi) (1 - e^{2\pi i\xi})^{-1} d\xi \\
 & + \omega \int_E^{x+E\omega} \phi(\theta) d\theta - \omega\phi(x) + c'',
 \end{aligned}$$

where c'' is an arbitrary constant. This is readily verified, and it is easy to see that conversely if the sequence R is chosen in any fashion so that, for a suitable choice of c' as a function of n , the limit $F(x; R)$ exists in \mathcal{J} , then $F(x; R)$ is given by (77). Hence (77) gives precisely the totality of $P(\text{Arg } \omega, \mathcal{J})$ solutions. But equation (77) is a step in Nörlund's work leading to equation (58). (Cf. "*Differenzenrechnung*" page 70, equation (9).)¹⁷ Hence the $P(\text{Arg } \omega, \mathcal{J})$ solutions coincide with the Nörlund principal solutions, for equation (1).

Section B. The $P(\arg \omega, \mathcal{J})$ solutions of equation (3).

The treatment is similar to the treatment of equation (1), but simpler, particularly because there is no arbitrary constant to consider. The $P(\text{Arg } \omega, \mathcal{J})$ solution, which is a unique function, (not a one-parameter family of functions), coincides with the Nörlund principal solution (59).

¹⁷ The arbitrary constant c'' , which has no obvious counterpart in Nörlund's equation (9), is innocuous because Nörlund's principal solution has a (somewhat concealed) additive arbitrary constant, the presence of which is indicated by the arbitrary α appearing in the cited equation (9).

PART VII. A Simple Linear Example Emphasizing the Influence of α upon the Principal Solution in Direction α .

THEOREM 6. *Given the difference equation*

$$(78) \quad y(x) - 2y(x+1) = \phi(x),$$

where $\phi(x)$ is analytic at ∞ . Let the half-plane $\Re(xe^{-i\alpha}) > D$ be denoted by $\mathcal{H}(\alpha, D)$. Let $\alpha_k = \text{Arg}(\text{Log } 2 + 2\pi ki)$, ($k = 0, \pm 1, \pm 2, \dots$). Then the $P(\alpha, \mathcal{H}(\alpha, D))$ solutions of (78) may be described as follows:

a. If $\alpha = \alpha_k$, the $P(\alpha, \mathcal{H}(\alpha, D))$ solution, for D sufficiently large, is of the form $F_k(x) + ce^{(\text{Log}(1/2) + 2\pi ki)x}$, where $F_k(x)$ is analytic and for some positive constant M satisfies the inequality $|F_k(x)| < M|x|$ in $\mathcal{H}(\alpha, D)$, and where c is an arbitrary constant.

b. If $\alpha \neq \alpha_k$, the $P(\alpha, \mathcal{H}(\alpha, D))$ solution, for D sufficiently large, is a uniquely determined function $F(x; \alpha)$, analytic and bounded in $\mathcal{H}(\alpha, D)$.

c. In all cases, every solution $y(x)$ of (78) which in a half-plane $\mathcal{H}(\alpha, D_1)$ is analytic and satisfies for some positive constants M_1, N the inequality $|y(x)| < M_1|x|^N$, is for some positive D_2 coincident in $\mathcal{H}(\alpha, D_2)$ with a $P(\alpha, \mathcal{H}(\alpha, D_2))$ solution.¹⁸

Proof. Let C be such that $\phi(x)$ is analytic in the set $|x| \geq C > 0$.

Case 1. Let α be different from every α_k , and let $\cos \alpha > 0$. Let $X = xe^{-i\alpha}$. Then, because of Theorem 1, a necessary and sufficient condition for $y_0(x)$ to be a $P(\alpha, \mathcal{H}(\alpha, D))$ solution of (78) is that the function $Y_0(X)$ defined by $Y_0(X) \equiv y_0(Xe^{i\alpha})$ be a $P(0, \mathcal{H}(0, D))$ solution of

$$(79) \quad Y_0(X) - 2Y_0(X + \omega) = \psi(X),$$

where $\omega = e^{-i\alpha}$ and $\psi(X) \equiv \phi(Xe^{i\alpha})$.

By means of Lemma V it is readily checked that the hypotheses of Theorem 3 are satisfied, so that the $P(0, \mathcal{H}(0, D))$ solution of (79) is unique and bounded, if D is sufficiently large, and therefore the $P(\alpha, \mathcal{H}(\alpha, D))$ solution of (78) is unique and bounded, if D is sufficiently large. Now let $y(x)$ be any solution of (78), analytic and satisfying a condition $|y(x)| < M_1|x|^N$ in a half-plane $\mathcal{H}(\alpha, D_1)$. Let $D_2 \geq D_1$, and $D_2 \geq D$. Let $F(x; \alpha)$ be the $P(\alpha, \mathcal{H}(\alpha, D))$ solution of (78). Then $y(x) - F(x; \alpha)$

¹⁸ The exponential functions $e^{(\text{Log}(1/2) + 2\pi ki)x}$ appearing in the $P(\alpha_k, \mathcal{H}(\alpha_k, D))$ solutions are bounded in $\mathcal{H}(\alpha_k, D)$.

is a solution of the homogeneous equation $y(x+1) - 2y(x) = 0$, and satisfies an inequality $|y(x) - F(x; \alpha)| < M_2 |x|^N$ in $\mathcal{H}(\alpha, D_2)$, whence by Lemma XXIV, $y(x) - F(x; \alpha) = 0$ in $\mathcal{H}(\alpha, D_2)$.

Case 2. Let $\cos \alpha < 0$. Let $X = xe^{-i\alpha}$. Then, because of Theorems 1 and 2, a necessary and sufficient condition for $y_0(x)$ to be a $P(\alpha, \mathcal{H}(\alpha, D))$ solution of (78) is that the function $Y_0(X)$ defined by $Y_0(X) \equiv y_0(Xe^{i\alpha} + 1)$ be a $P(0, \mathcal{H}(0, D - \cos \alpha))$ solution of

$$(80) \quad Y_0(X + \omega) - 2Y_0(X) = \psi(X),$$

where $\omega = -e^{-i\alpha}$ and $\psi(X) = \phi(Xe^{i\alpha})$. We now apply Theorem 3, and Lemma XXIV, as in Case 1.

Case 3. Let $\cos \alpha = 0$. (This case cannot be brought under Theorem 3, in the manner used for Cases 1 and 2, since after such rotation of the variable we do not have the condition $\Re(\omega) > 0$ satisfied.)

We note that if q_t , ($t = 1, 2, \dots$), is a sequence of complex numbers, of positive imaginary part, and of modulus unity, and if the sequence tends to unity as a limit, then numbers b_t , ($t = 1, 2, \dots$), can be found such that (q_t, b_t) , ($t = 1, 2, \dots$), will be a $Q(\pi/2)$ sequence, and $(1/q_t, -b_t)$, ($t = 1, 2, \dots$) will be a $Q(-\pi/2)$ sequence. (Cf. Lemmas II and VI.)

Let

$$(81) \quad z_0 y(x) - 2z_1 y(x+1) = \phi(x)$$

be the parametrized equation for (78). Let (q_t, b_t) , ($t = 1, 2, \dots$), be a $Q(\alpha)$ -sequence chosen to satisfy the additional condition $|q_t| \geq 1$. Let

$$(82) \quad z_0 y(x) - 2z_0 y(qx + V) = \phi(x)$$

be the corresponding approximating q -difference equation. Let $y(x; t; z_0, z_1)$ be a special solution of (82). Then if $\phi(x) = \sum_{k=0}^{\infty} \phi_k u^k$, where $u = x - b$, evidently

$$(83) \quad y(x; t; z_0, z_1) = \sum_{k=0}^{\infty} \phi_k u^k / (z_0 - 2z_1 q^k).$$

If z_0, z_1 are both near unity, then $|z_0 - 2z_1 q^k| > 1/2$, since $|q^k| \geq 1$, ($k = 0, 1, \dots$). Hence

$$\begin{aligned} |y(x; t; z_0, z_1)| &\leq 2 \sum_{k=0}^{\infty} |\phi_k| |u^k| \leq 2 \sum_{k=0}^{\infty} |u|^k MC / (|b| - C)^{k+1} \\ &= 2MC / (|b| - C - |u|). \end{aligned}$$

(Cf. Lemma XVI.) Thus, if $|x - b| < |b| - C_1$, with $C_1 > C$, we have $|y(x; t; z_0, z_1)| < 2MC/(C_1 - C)$, from which, by the compactness theorem for bounded families of analytic functions, we conclude the existence of at least one $Q(\alpha, \mathcal{H}(\alpha, C))$ solution $y(x; z_0, z_1)$ of (81), for all z_0, z_1 such that $2|z_1| - |z_0| > 1/2$, bounded in every half-plane $\mathcal{H}(\alpha, C_1)$ with $C_1 > C$, and also conclude the existence of at least one $P(\alpha, \mathcal{H}(\alpha, C))$ solution $y(x)$ of (78), likewise bounded in every half-plane $\mathcal{H}(\alpha, C_1)$ with $C_1 > C$.

We wish to prove next that every $Q(\alpha, \mathcal{H}(\alpha, D))$ solution of (81), for $D > C$, is majorized in $\mathcal{H}(\alpha, D)$ by $2MC/(D - C)$. Let $y^*(x; z_0, z_1)$ be for some positive D a $Q(\alpha, \mathcal{H}(\alpha, D))$ solution of (81). Let (q_t, b_t) , $(t = 1, 2, \dots)$, be a $Q(\alpha)$ sequence, and let $y(x; t; z_0, z_1)$ be a special solution of (82), such that $y^*(x; z_0, z_1) = \lim_{t \rightarrow \infty} y(x; t; z_0, z_1)$, the limit being uniform in every closed bounded subset of $\mathcal{H}(\alpha, D)$. By virtue of the discussion just above, for the case $|q_t| \geq 1$, we may and do confine our attention to the case where $|q_t| < 1$ for every t .

Let $\lambda_0 = \lambda_0(t)$ be the positive number such that $|z_0| - 2|z_1| + |q|^\lambda = 0$. Then

$$(84) \quad y(x; t; z_0, z_1) = \sum' \phi_k u^k (z_0 - 2z_1 q^k)^{-1} + c u^{\lambda_0},$$

where \sum' is the sum over all non-negative integers different from λ_0 , and c is a function of t , constant with respect to x .

Let $\psi_t = \text{Arg } q_t$, $L_t = \text{Log } |q_t|$. Then $\lim_{t \rightarrow \infty} L_t/\psi_t = 0$, by Lemma II. Let $\epsilon_t = L_t/\psi_t$.

Let $\sigma_t = e^{i\psi_t}$. Then (σ_t, b_t) , $(t = 1, 2, \dots)$, is a $Q(\alpha)$ sequence. Let $h(x; t; z_0, z_1)$ be a special solution of

$$(85) \quad z_0 h(x) - 2z_1 h(\sigma x + W) = \phi(x),$$

where $W = b(1 - \sigma)$. Then for some subsequence J of the positive integers, (which we may and do suppose to be the entire sequence of positive integers), $h(x; t; z_0, z_1)$ approaches a limit function $h(x; z_0, z_1)$ as t becomes infinite on J , the limit being uniform in every closed bounded subset of $\mathcal{H}(\alpha, C)$; moreover, $|h(x; z_0, z_1)| < 2MC/(C_1 - C)$ in every half-plane $\mathcal{H}(\alpha, C_1)$, with $C_1 > C$. (All this follows, since $|\sigma| = 1$, from the earlier discussion where $|q_t| \geq 1$.) Now

$$(86) \quad h(x; t; z_0, z_1) = \sum_{k=0}^{\infty} \phi_k u^k (z_0 - 2z_1 \sigma^k)^{-1},$$

so that

$$(87) \quad \begin{aligned} y(x; t; z_0, z_1) - h(x; t; z_0, z_1) \\ = 2z_1 \sum' \phi_k u^k (q^k - \sigma^k) (z_0 - 2z_1 q^k)^{-1} (z_0 - 2z_1 \sigma^k)^{-1} + c_1 u^{\lambda_0} \end{aligned}$$

for some constant c_1 , (depending on t). We take $C_1 > C$, and $D_4 > D_3 > C_1$, with $D_4 - D_3 < (C_1 - C)/2$, take x_0 so that $D_3 < \Re(x_0 e^{-ia}) < D_4$, and define $u_0 = x_0 - b$. Then

$$(88) \quad y(x; t; z_0, z_1) - h(x; t; z_0, z_1) \\ = 2z_1 u^{\lambda_0} \Sigma' \phi_k (q^k - \sigma^k) (z_0 - 2z_1 q^k)^{-1} (z_0 - 2z_1 \sigma^k)^{-1} (u^{k-\lambda_0} - u^{k-\lambda_0}) \\ + c_2 u^{\lambda_0},$$

for some constant c_2 , (depending upon t). Let $H(x; t; z_0, z_1)$ be the right-hand member of (88), deprived of the term $c_2 u^{\lambda_0}$. Then it is easy to see that if $D_3 < \Re(x e^{-ia}) < D_4$, then $|H(x; t; z_0, z_1)| < M_1 |x - x_0|$ for some positive M_1 . (To verify this we note that (a) if $k < \lambda_0/2$, then

$$|z_0 - 2z_1 q^k| > A > 0, \quad |q^k - \sigma^k| < 2, \quad |z_0 - 2z_1 \sigma^k| > A_1 > 0. \\ |u|^{\lambda_0} |u^{k-\lambda_0} - u_0^{k-\lambda_0}| < 2(|b| - C_1)^k, \quad \text{and} \quad |\phi_k| < MC/(|b| - C)^{k+1},$$

while (b) if $\lambda_0/2 \leq k \leq \lambda_0 + 1$, then

$$|z_0 - 2z_1 q^k| > A_2 |q|^{2k} |\lambda_0 - k| (1 - |q|), \quad (\text{with } A_2 > 0), \\ |q^k - \sigma^k| < k |q - \sigma| = k(1 - |q|), \quad |z_0 - 2z_1 \sigma^k| > A_1 > 0, \\ |u^{\lambda_0}| < (|b| - D_3)^{\lambda_0},$$

$|u^{k-\lambda_0} - u_0^{k-\lambda_0}| < |\lambda_0 - k| |x - x_0| (|b| - D_4)^{k-1-\lambda_0}$, and finally (c) if

$$\lambda_0 + 1 < k, \quad \text{then} \quad |u^{\lambda_0}| < (|b| - D_3)^{\lambda_0}, \quad |q^k - \sigma^k| < k(1 - |q|), \\ |z_0 - 2z_1 q^k| > A_3 (1 - |q|)(k - \lambda_0), \quad |z_0 - 2z_1 \sigma^k| > A_4, \\ |u^{k-\lambda_0} - u_0^{k-\lambda_0}| < (|b| - D_3)^{k-\lambda_0-1} |x - x_0| (k - \lambda_0).$$

Hence for some subsequence J_1 of the positive integers, limit $H(x; t; z_0, z_1)$ $\xrightarrow[t \in J_1]{t \rightarrow \infty}$ exists in $D_3 < \Re(x e^{-ia}) < D_4$, uniformly in every closed bounded subset of the strip. It follows that in that strip $c_2 u^{\lambda_0}$ approaches a limit as t becomes infinite on J_1 , uniformly in every closed bounded subset of that strip. By Lemma XXIII this limit of $c_2 u^{\lambda_0}$ must be identically zero in the strip, since λ_0/b is easily seen to become infinite with t . Thus $|y(x; z_0, z_1) - h(x; z_0, z_1)| < M_2 |x - x_0|$ in the strip $D_3 < \Re(x e^{-ia}) < D_4$. But this implies that $y^*(x; z_0, z_1) - h(x; z_0, z_1) \equiv 0$, since this difference is a solution of the homogeneous equation $z_0 y(x) - 2z_1 y(x+1) = 0$, and no solution of this

which is not identically zero can be majorized by $M_2 |x - x_0|$ in the (horizontal) strip $D_3 < \Re(xe^{-ia}) < D_4$. Hence $h(x; z_0, z_1)$ coincides with $y(x; z_0, z_1)$ in the half-plane $\mathcal{H}(\alpha, D)$, and thus $|h(x; z_0, z_1)| < MC/(D - C)$ in $\mathcal{H}(\alpha, D)$.

Plainly it follows from this discussion that every $P(\alpha, \mathcal{H}(\alpha, D))$ solution of (78) is majorized by $MC/(D - C)$ in $\mathcal{H}(\alpha, D)$, if $D > C$. The rest of the theorem for this case now follows immediately from Lemma XXIV.

Case 4. $\alpha = \alpha_k$. We first find a particular $P(\alpha, \mathcal{H}(\alpha, C))$ solution of (78). Let q_t , ($t = 1, 2, \dots$), be defined by the equations $\text{Log } |q_t| = (t + (1/2))^{-1} \text{Log } (1/2)$, $\text{Arg } q_t = 2\pi k/(t + (1/2))$. Let $b_t = (1 - q_t)^{-1}$, ($t = 1, 2, \dots$). Then (q_t, b_t) , ($t = 1, 2, \dots$), is a $Q(\alpha)$ -sequence. We note also that $\lim_{t \rightarrow \infty} |b_t| |\text{Log } |q_t|| = A_k$, where

$$A_k = (1 + (2\pi k)^2 (\text{Log } 2)^{-2})^{-(1/2)}.$$

Let $y(x, t)$ be a special solution of

$$(89) \quad y(x) - 2y(qx + 1) = \phi(x).$$

Then $y(x, t) = \sum_{k=0}^{\infty} \phi_k u^k (1 - 2q^k)^{-1} + cu^\lambda$, where c is an arbitrary constant

and $\lambda = t + (1/2)$. Hence $y(x, t) = u^\lambda \sum_{k=0}^{\infty} \phi_k (u^{k-\lambda} - u_0^{k-\lambda}) (1 - 2q^k)^{-1} + du^\lambda$, where d is an arbitrary constant. Let $y_1(x, t) = y(x, t) - du^\lambda$. Using estimates similar to those noted in Case 3, we verify easily that for every choice of positive numbers D_1 and D_2 , with $D_2 > D_1 > C$, there is a number $M(D_1, D_2)$ such that if \mathcal{F} is a closed bounded set included in the strip $D_2 > \Re(xe^{-ia}) > D_1$, then when t is large $|y_1(x, t)| < M(D_1, D_2) |x|$ for all x in \mathcal{F} . Hence, by the use of $d = 0$, a $Q(\alpha, \mathcal{H}(\alpha, C))$ solution $y_0(x)$ of (78) is obtainable, such that $|y_0(x)| < M(D_1, D_2) |x|$ in every strip $D_2 > \Re(xe^{-ia}) > D_1 > C$. Of course $y_0(x)$ is also a $P(\alpha, \mathcal{H}(\alpha, C))$ solution of (78). By using the arbitrariness of d , we readily obtain

$$y_0(x) + c_1 e^{x(\text{Log } (1/2) + 2\pi k i)}$$

as a $P(\alpha, \mathcal{H}(\alpha, C))$ solution of (78), with c_1 arbitrary.

Now since $|y_0(x)| \leq M(D_1, D_2) |x|$ in every strip $D_2 > \Re(xe^{-ia}) > D_1 > C$, it follows from equation (78) itself that for every D greater than C there is a positive number $M(D)$ such that $|y_0(x)| \leq M(D) |x|$ in the half-plane $\mathcal{H}(\alpha, D)$. (We use the relation, obvious from (78), that

$$y_0(x + n) = h_0(x) 2^{-n} + (-1) \sum_{k=0}^{n-1} \phi(x + k) 2^{n-k}.)$$

Next we consider any choice of z_0, z_1 near 1, 1, and any $Q(\alpha)$ -sequence, and following steps similar to those used in obtaining $y_0(x)$, (the essential difference in treatment being that for some choices of z_0, z_1 , and some choices of the $Q(\alpha)$ -sequence the term with arbitrary coefficient may be lacking), we prove that every $Q(\alpha, \mathcal{H}(\alpha, D))$ solution of (81) for $D > C$, must be of the form $y_1(x) + c''e^{x(\text{Log}(z_0/2z_1) + 2\pi ki)}$, where $y_1(x)$ is majorized by an expression of the form $M_2 |x|$ in $\mathcal{H}(\alpha, D)$, M_2 being independent of z_0, z_1 .

From this it follows that every $P(\alpha, \mathcal{H}(\alpha, D))$ solution of (78) is of the form $y_2(x) + c'''e^{x(\text{Log}(\frac{1}{2}) + 2\pi ki)}$, where $y_2(x)$ is majorized by $M_2 |x|$ in $\mathcal{H}(\alpha, D)$.

The remaining conclusions of the theorem then follow, in Case 4, from Lemma XXIV.

PART VIII. Appendix.

Section A. The fundamental relation between the expansion of $F(x)$ and the expansion of $F(\sigma x + V)$.

LEMMA I. If $F(x) = \sum_{\lambda} F_{\lambda}(x-b)^{\lambda}$ is a generalized power-series at $x=b$, convergent in a quasi-neighborhood \mathcal{N} of b , and if σ, V are complex numbers such that $b(1-\sigma) = V$, then

$$(90) \quad F(\sigma x + V) = \sum_{\lambda} F_{\lambda} \sigma^{\lambda} (x-b)^{\lambda},$$

(where $\sigma^{\lambda} = e^{\lambda \text{Log } \sigma}$), in the quasi-neighborhood \mathcal{N}' consisting of all points x such that $\sigma x + V$ is in \mathcal{N} .

Proof. Obvious.

Section B. $Q(\alpha)$ -sequences.

LEMMA II. Let q_1, q_2, \dots be a sequence of complex numbers, each distinct from unity, the limit of the sequence being unity. Let α be a number such that $-\pi < \alpha \leq \pi$. Then a necessary and sufficient condition that there exist a sequence b_1, b_2, \dots such that the sequence of pairs (q_t, b_t) , ($t = 1, 2, \dots$), be a $Q(\alpha)$ -sequence is that

$$(91) \quad \lim_{t \rightarrow \infty} (\text{Arg } (q_t)) / |\text{Log } q_t| = \sin \alpha,$$

and

$$(92) \quad \lim_{t \rightarrow \infty} (\text{Log } |q_t|) / |\text{Log } q_t| = -\cos \alpha.$$

Proof. Necessity. Since (q_t, b_t) , $(t = 1, 2, \dots)$, is a $Q(\alpha)$ -sequence, we must have $b_t \text{Log } q_t = -\theta(t)$, where $\theta(t)$ tends to unity as t becomes infinite. Hence $\text{Arg } \text{Log } q_t \equiv \pi + \text{Arg } \theta(t) - \alpha + \eta(t) \pmod{2\pi}$, where $\eta(t)$ tends to zero as t becomes infinite. This implies $\text{Arg } q_t / |\text{Log } q_t| = \sin(\pi + \text{Arg } \theta(t) - \alpha + \eta(t))$ and $\text{Log } |q_t| / |\text{Log } q_t| = \cos(\pi + \text{Arg } \theta(t) - \alpha + \eta(t))$, from which the necessity follows at once.

Sufficiency. We define b_t by the equation $b_t = |\text{Log } q_t|^{-1} e^{i\alpha}$. Then $b_t \text{Log } q_t = e^{i(\alpha + \text{Arg } \text{Log } q_t)}$, and since from the hypotheses $\text{Arg } \text{Log } q_t \equiv \pi - \alpha + \eta(t) \pmod{2\pi}$, with $\eta(t)$ tending to zero as t becomes infinite, we have $b_t (\text{Log } q_t)$ tending to -1 as t becomes infinite. Since $\lim_{t \rightarrow \infty} (\text{Log } q_t) / (1 - q_t) = -1$, this implies that $\lim_{t \rightarrow \infty} b_t (1 - q_t) = 1$.

LEMMA III. Let (q_t, b_t) , $(t = 1, 2, \dots)$, be a $Q(\alpha)$ -sequence. Let ω be a non-zero complex number. Let $\sigma_t = q_t^\omega$, where $q_t^\omega = e^{\omega \text{Log } q_t}$, $(t = 1, 2, \dots)$. Then

$$(93) \quad \lim_{t \rightarrow \infty} (\text{Arg } \sigma_t) / |\text{Log } \sigma_t| = \sin(\alpha - \text{Arg } \omega),$$

and

$$(94) \quad \lim_{t \rightarrow \infty} (\text{Log } |\sigma_t|) / |\text{Log } \sigma_t| = -\cos(\alpha - \text{Arg } \omega).$$

Proof. Since $\sigma_t = e^{\omega \text{Log } q_t}$, we have $\text{Log } \sigma_t \equiv \omega \text{Log } q_t \pmod{2\pi}$, and since $\text{Log } q_t$ tends to zero as t becomes infinite, this implies that if t is large then $\text{Log } \sigma_t = \omega \text{Log } q_t$. From this relation, and Lemma II, the theorem follows immediately.

LEMMA IV. If (q_t, b_t) , $(t = 1, 2, \dots)$, is a $Q(\alpha)$ -sequence, and ω is a complex number such that $\cos(\text{Arg } \omega - \alpha) > 0$, then $|q_t^\omega| < 1$ if t is large, and if ω is a complex number such that $\cos(\text{Arg } \omega - \alpha) < 0$, then $|q_t^\omega| > 1$ if t is large.

Proof. This follows immediately from Lemma III, equation (94).

LEMMA V. Let (q_t, b_t) , $(t = 1, 2, \dots)$, be a $Q(\alpha)$ -sequence, let ω be a non-zero complex number such that $\cos(\text{Arg } \omega - \alpha) \neq 0$, and let ξ be a non-zero complex number. Then a necessary and sufficient condition that there exist for every t a non-negative λ_t such that $q_t^{\omega\lambda} = \xi$, (where $q_t^{\omega\lambda} = e^{\omega\lambda \text{Log } q_t}$), is that either $\xi = 1$, or else for some integer k_0

$$(95) \quad [\text{Arg } \xi + 2\pi k_0] / |\text{Log } \xi + 2\pi k_0 i| = \sin(\text{Arg } \omega - 2\alpha),$$

and

$$(96) \quad [\operatorname{Log} |\xi|]/|\operatorname{Log} \xi + 2\pi k_0 i| = -\cos(\operatorname{Arg} \omega - \alpha),$$

and for every sufficiently large t .

$$(97) \quad \operatorname{Arg} \operatorname{Log} (q^\omega) \equiv \pi + \operatorname{Arg} \omega - \alpha \pmod{2\pi}.$$

Proof. Let $\sigma = q^\omega = e^{\omega \operatorname{Log} q}$. Then $\sigma^\lambda = e^{\lambda \operatorname{Log} \sigma} = \xi$. Hence $\lambda \operatorname{Log} \sigma = \operatorname{Log} \xi + 2\pi k(t)i$ for some integer $k(t)$. Thus, if $\lambda \neq 0$, then

$$(98) \quad [\operatorname{Arg} \xi + 2\pi k(t)]/|\operatorname{Log} \xi + 2\pi k(t)i| = (\operatorname{Arg} \sigma)/|\operatorname{Log} \sigma|,$$

and

$$(99) \quad (\operatorname{Log} |\xi|)/|\operatorname{Log} \xi + 2\pi k(t)i| = (\operatorname{Log} |\sigma|)/(|\operatorname{Log} \sigma|).$$

By (98), (99), and Lemma III, we have

$$(100) \quad \lim_{t \rightarrow \infty} [(\operatorname{Arg} \xi + 2\pi k(t))/|\operatorname{Log} \xi + 2\pi k(t)i|] = \sin(\alpha - \operatorname{Arg} \omega)$$

$$(101) \quad \lim_{t \rightarrow \infty} [(\operatorname{Log} |\xi|)/|\operatorname{Log} \xi + 2\pi k(t)i|] = -\cos(\alpha - \operatorname{Arg} \omega).$$

From (101) it follows that $k(t)$ is a constant k_0 for t large. Hence (95) and (96) hold.

Then by (98) and (99) $(\operatorname{Arg} \sigma)/|\operatorname{Log} \sigma| = \sin(\alpha - \operatorname{Arg} \omega)$ and $(\operatorname{Log} |\sigma|)/|\operatorname{Log} \sigma| = -\cos(\alpha - \operatorname{Arg} \omega)$, for all sufficiently large t , and these equations imply (97).

LEMMA VI. If (q_t, b_t) , $(t=1, 2, \dots)$, is a $Q(\alpha)$ -sequence, and ω is a non-zero complex number, then $(q_t^\omega, \omega^{-1}b_t)$ is a $Q(\beta)$ -sequence, where $\beta \equiv \alpha - \operatorname{Arg} \omega \pmod{2\pi}$.

Proof. This follows at once from Lemmas II and III, and the easily verified relation $\lim_{t \rightarrow \infty} [(\omega^{-1}b_t)(1 - q_t^\omega)] = 1$.

LEMMA VII. Let R_1, R_2, \dots be a sequence of numbers such that $0 < R_t < 1$ and $\lim_{t \rightarrow \infty} R_t = 1$. Let ω be a non-zero complex number. Let $q_t = R_t^{1/\omega} \equiv e^{(1/\omega) \operatorname{Log} R_t}$, $(t=1, 2, \dots)$. Then there is a sequence b_1, b_2, \dots such that the sequence of pairs (q_t, b_t) is a $Q(\operatorname{Arg} \omega)$ -sequence.

Proof. Evidently $(R_t, (1 - R_t)^{-1})$, $(t=1, 2, \dots)$, is a $Q(0)$ -sequence. Hence, by Lemma VI, $(R_t^{1/\omega}, (1/\omega)(1 - R_t)^{-1})$, $(t=1, 2, \dots)$ is a $Q(\beta)$ -sequence, where $\beta \equiv 0 - \operatorname{Arg}(1/\omega) \pmod{2\pi} \equiv \operatorname{Arg} \omega \pmod{2\pi}$.

Section C. Logarithmic spirals.

LEMMA VIII. Let ω be a non-zero complex number. Let (q_t, b_t) ,

($t = 1, 2, \dots$), be a $Q(\text{Arg } \omega)$ -sequence. Let β be a number such that $0 \leq \beta < (\pi/2)$. Let $\xi_1(\rho) = \rho e^{i\beta}$, $\xi_2(\rho) = \rho e^{-i\beta}$ ($0 < \rho < \infty$). Then if t is sufficiently large we shall have $|q^{\omega \xi_1(\rho)}| \leq 1$, $|q^{\omega \xi_2(\rho)}| \leq 1$ for all non-negative ρ .

Proof. Let $\sigma_t = q_t^\omega$. Then $q^{\omega \xi_j(\rho)} = \sigma^{\xi_j(\rho)} = e^{\xi_j \text{Log } \sigma}$ ($j = 1, 2$). Hence

$$\begin{aligned} (102) \quad |q^{\omega \xi_j(\rho)}| &= \exp\{\Re(\xi_j \text{Log } \sigma)\} \\ &= \exp\{\rho[(\cos \beta) \text{Log } |\sigma| - (\sin \beta) \text{Arg } \sigma]\} \\ &= \exp\{\rho \cos \beta \text{Log } |\sigma| (1 - \tan \beta (\text{Arg } \sigma) / (\text{Log } |\sigma|))\}. \end{aligned}$$

By Lemma III, with $\alpha = \text{Arg } \omega$, limit $[(\text{Arg } \sigma) / (\text{Log } |\sigma|)] = 0$. Hence $1 - \tan \beta (\text{Arg } \sigma) / (\text{Log } |\sigma|) > 0$ if t is large. Also, $\text{Log } |\sigma| < 0$ by Lemma IV. Thus $|q^{\omega \xi_j(\rho)}| \leq e^0 = 1$.

LEMMA IX. If $\Re(y) < \eta$, then $|(1 - e^y)/y| \leq e^{|\eta|}$.

Proof. $(e^y - 1)/y = \int_0^1 e^{yt} dt$, where the path of integration is a straight line segment. Hence $|(e^y - 1)/y| \leq \max_{0 \leq t \leq 1} |e^{yt}| = \max_{0 \leq t \leq 1} \exp[\Re(yt)] = \max_{0 \leq t \leq 1} \exp[t \Re(y)] \leq e^{|\eta|}$.

LEMMA X. Under the hypotheses of Lemma VIII, let $\xi_j = E + \zeta_j(\rho)$, ($j = 1, 2$), where $0 < E < 1$. Then for every positive δ there is a positive T such that if $t > T$, then

$$|q^{\omega \xi_j}(x - b) + b| < |x| + (|\omega| + \delta)|\xi_j|, \quad (j = 1, 2).$$

Proof. $q^{\omega \xi_j}(x - b) + b = q^{\omega \xi_j}x + b(1 - q^{\omega \xi_j})$.

Now $|q^{\omega \xi_j}| = |q^{\omega E}| |q^{\omega \zeta_j(\rho)}|$ and each factor is smaller than unity, if t is large (by Lemmas IV and VIII). Thus $|q^{\omega \xi_j}x| \leq |x|$, ($j = 1, 2$), if t is large.

Let $V = b(1 - q^\omega)$. Let δ be any positive number. Let η be any positive number, to be specified later in terms of δ . Let T be so large that if $t > T$, then $|V - \omega| < \eta$, and

$$|[(\omega \text{Log } q) / (e^{\omega \text{Log } q} - 1)] - 1| < \eta,$$

and

$$|q^{\omega \xi_j(\rho)}| \leq 1, \quad (j = 1, 2),$$

for all negative ρ . (See Lemma VIII.) Then $|q^{\omega \xi_j}| \leq 1$, so that

$$\Re(\omega \xi_j \text{Log } q) \leq 0,$$

and therefore by Lemma IX we have

$$|(1 - e^{\omega \xi_j \text{Log } q}) / (\omega \xi_j \text{Log } q)| \leq 1.$$

Now

$$\begin{aligned} (103) \quad b(1 - q^{\omega \xi_j}) &= V \frac{(1 - q^{\omega \xi_j})}{(1 - q^\omega)} \\ &= (V \xi_j) \cdot \frac{e^{\omega \xi_j \text{Log } q} - 1}{\omega \xi_j \text{Log } q} \cdot \frac{\omega \text{Log } q}{e^{\omega \text{Log } q} - 1}. \end{aligned}$$

Hence $|b(1 - q^{\omega \xi_j})| < (|\omega| + \eta)|\xi_j|(1 + \eta)$. If η is sufficiently small, these results imply

$$|q^{\omega \xi_j}(x - b) + b| < |x| + (|\omega| + \delta)|\xi_j|, \quad (j=1, 2).$$

LEMMA XI. Let $\mathcal{B}_1, \mathcal{B}_2$ be any bounded sets in the x, ξ planes, respectively. Let (q_t, b_t) be a $Q(\alpha)$ -sequence. Let ω be a complex number. For every positive δ there exists a positive T such that if $t > T$, then

$$| [q^{\omega \xi}(x - b) + b] - [x + \omega \xi] | < \delta$$

for all x in \mathcal{B}_1 and all ξ in \mathcal{B}_2 .

Proof. Let $V = b(1 - q^\omega)$. Then

$$\begin{aligned} (104) \quad | [q^{\omega \xi}(x - b) + b] - [x + \omega \xi] | &= \left| x(e^{\omega \xi \text{Log } q} - 1) + (V - \omega)\xi \right. \\ &\quad \left. + (V\xi) \left(\frac{e^{\omega \xi \text{Log } q} - 1}{\omega \xi \text{Log } q} - 1 \right) + V\xi \left(\frac{e^{\omega \xi \text{Log } q} - 1}{\omega \xi \text{Log } q} \right) \left(\frac{\omega \text{Log } q}{e^{\omega \text{Log } q} - 1} - 1 \right) \right|. \end{aligned}$$

Let $|x| < M_1$ in \mathcal{B}_1 , $|\xi| < M_2$ in \mathcal{B}_2 . Let $\delta > 0$. Let $\eta > 0$ be such that $|e^\gamma - 1| < \delta/(4M_1)$ if $|\gamma| < \eta$, and such that

$$\left| \frac{e^\gamma - 1}{\gamma} - 1 \right| < \delta/(4M_2(|\omega| + 1))$$

if $|\gamma| < \eta$, and such that

$$\left| \frac{e^{\gamma_1} - 1}{\gamma_1} \right| \left| \frac{\gamma_2}{e^{\gamma_2} - 1} - 1 \right| < \delta/(4M_2(|\omega| + 1))$$

if $|\gamma_1| < \eta$ and $|\gamma_2| < \eta$. Let T be such that if $t > T$ then all the inequalities $|V - \omega| < 1$, $|V - \omega| < \delta/(4M_2)$, $|\text{Log } q| < \eta/(M_2|\omega|)$, and $|\text{Log } q| < \eta/|\omega|$ are valid.

Then if $t > T$ we have

$$\begin{aligned} | [q^{\omega \xi}(x - b) + b] - [x + \omega \xi] | &< M_1[\delta/(4M_1)] + [\delta/(4M_2)]M_2 \\ &\quad + (|\omega| + 1)M_2[\delta/(4M_2(|\omega| + 1))] \\ &\quad + (|\omega| + 1)M_2[\delta/(4M_2(|\omega| + 1))] \\ &= \delta. \end{aligned}$$

LEMMA XII. (1) Let $f(X) = Pe^{X \cot C} - \sin X$ where $P > 0$ and $0 < C < \pi/2$. Then $f(X) > 0$ throughout the interval (X_1, ∞) , where X_1 satisfies $0 < X_1 < \pi$, provided all the following are valid:

- a. $P > \sin X_1 e^{-X_1 \cot C}$
- b. $X_1 > C$
- c. $P > e^{-(2\pi+C) \cot C} \sin C$.

(2) Let $g(Y) = Re^{Y \cot C} + \sin Y$, where $R > 0$ and $0 < C < \pi/2$. Then $g(Y) > 0$ throughout (Y_1, ∞) , where Y_1 satisfies $-\pi < Y_1 < 0$, provided all the following are valid:

- a. $R > |\sin Y_1| e^{-Y_1 \cot C}$
- b. $Y_1 > C - \pi$
- c. $R > e^{-(\pi+C) \cot C} \sin C$.

Proof by elementary calculus.

LEMMA XIII. Let \mathcal{S} be the sector $-\pi/2 \leq M < \arg \zeta < L \leq \pi/2$. Let ω be a point of \mathcal{S} . Let $\alpha = \text{Arg } \omega$. Let $\beta_0 = \text{minimum } (\pi/2, L - \alpha, \alpha - M)$. Let $0 < \beta < \beta_0$. Let \mathcal{F} be a closed bounded subset of \mathcal{S} . Let (q_t, b_t) , $(t = 1, 2, \dots)$, be a $Q(\alpha)$ -sequence. Let $\xi_1 = q^{wz_1}(x - b) + b$, $\xi_2 = q^{wz_2}(x - b) + b$, with $z_1 = \rho(\cos \beta + i \sin \beta)$, $z_2 = \rho(\cos \beta - i \sin \beta)$ and $0 \leq \rho \leq \infty$. Then if t is sufficiently large, ξ_1 and ξ_2 are both in \mathcal{S} for every non-negative ρ and every x of \mathcal{F} .

Proof. We shall consider only ξ_1 , since ξ_2 is treated similarly. Let $\sigma_t = q_t \omega$, $(t = 1, 2, \dots)$. Then $(\sigma_t, \omega^{-1}b_t)$, $(t = 1, 2, \dots)$, is a $Q(0)$ -sequence. (By Lemma VI). Hence, by Lemma II,

$(\text{Arg } \sigma_t)/|\text{Log } \sigma_t| = \sin \eta_t$, and $(\text{Log } |\sigma_t|)/|\text{Log } \sigma_t| = -\cos \eta_t$, where $\lim_{t \rightarrow \infty} \eta_t = 0$. Let $\delta = \delta_t(x) = \text{Log } |1 - x/b_t|$, and let $\epsilon = \epsilon_t(x) = \text{Arg } (1 - x/b_t)$. It is readily seen that

$$(105) \quad \Re(\xi_1/b) = 1 - e^{-r \cos(\beta - \eta) + \delta} \cos(r \sin(\beta - \eta) - \epsilon)$$

and

$$(106) \quad \Im(\xi_1/b) = e^{-r \cos(\beta - \eta) + \delta} \sin(r \sin(\beta - \eta) - \epsilon)$$

where $r = \rho |\text{Log } \sigma_t|$. Thus r varies from 0 to $+\infty$ as ρ varies from 0 to $+\infty$. Let $M_1 = M - \text{Arg } b$, $L_1 = L - \text{Arg } b$. We are to prove that

$$(107) \quad M_1 < \text{Arg } (\xi_1/b) < L_1.$$

If r is large, $\Re(\xi_1/b)$ is near 1 and $\Im(\xi_1/b)$ is near 0. Hence (107) is satisfied if r is large. Hence if (107) fails for some non-negative r , then for some non-negative r either

$$(108) \quad \text{Arg}(\xi_1/b) = L_1$$

or

$$(109) \quad \text{Arg}(\xi_1/b) = M_1.$$

By elementary calculations it is found that (105), (106) and (108) imply that the equation

$$(110) \quad Pe^{X \cot C} - \sin X = 0$$

holds for some X in (X_1, ∞) , where

$$X_1 = L_1 - \epsilon, \quad C = \beta - \eta, \quad X = L_1 + r \sin C - \epsilon,$$

and

$$P = \sin L_1 e^{-(L_1 - \epsilon) \cot C - \delta},$$

and that (105), (106) and (109) imply that the equation

$$(111) \quad Re^{Y \cot C} + \sin Y = 0$$

holds for some Y in (Y_1, ∞) , where

$$Y_1 = M_1 - \epsilon, \quad C = \beta - \eta, \quad Y = M_1 + r \sin C - \epsilon,$$

and

$$R = |\sin M_1| e^{-(M_1 - \epsilon) \cot C - \delta}.$$

It is readily seen that (110) contradicts Lemma XII (1), while (111) contradicts Lemma XII (2).

Section D. Almost constant functions.

Definition I. Let $S: (q_t, b_t)$, $(t=1, 2, \dots)$, be a $Q(\alpha)$ -sequence. Let $f(x, t)$ be a function of x and t , which is defined for every sufficiently large positive integer t as a function of x analytic in a neighborhood of b_t , with Taylor's expansion $\sum_{k=0}^{\infty} f_k(t) (x - b_t)^k$. The function $f(x, t)$ will be called "*almost constant on the sequence S*" if $f_0(t)$ is a bounded function of t , and if for every positive ϵ there is a positive D , (independent of t), such that for all sufficiently large t the inequality

$$(112) \quad \sum_{k=1}^{\infty} |f_k(t)| (|b_t| - D)^k < \epsilon$$

is valid.

Definition II. Let $F(x, b)$ be a function of x and b which, for all sufficiently large b such that $\text{Arg } b = \alpha$, is analytic at $x = b$, with Taylor's expansion $\sum_{k=0}^{\infty} F_k(b)(x - b)^k$. We shall say that $F(x, b)$ is "almost constant in direction α " if $F(b, b)$ is a bounded function of b , when $\text{Arg } b = \alpha$, and b is large, and if for every positive ϵ there is a positive D (independent of b), such that for all sufficiently large b the inequality

$$(113) \quad \sum_{k=1}^{\infty} |F_k(b)| (|b_t| - D)^k < \epsilon$$

is valid.

LEMMA XIV. Let $F(x, b)$ be almost constant in direction α . Let $S: (q_t, b_t)$, $(t = 1, 2, \dots)$, be a $Q(\alpha)$ -sequence. Let $f(x, t)$ be defined as the function $F(x, |b_t| e^{i\alpha})$. Then $f(x, t)$ is almost constant on S .

Proof. Let $F(x, c) = \sum_{k=0}^{\infty} F_k(c)(x - c)^k$ for all c such that $\text{Arg } c = \alpha$. Let $c_t = |b_t| e^{i\alpha}$. Then $|c_t - b_t| = |b_t| |e^{i\alpha} - e^{i \text{Arg } b}| = O(1)$. Let ϵ be a positive number. Let D_1 be such that

$$\sum_{k=1}^{\infty} |F_k(c)| (|c| - D_1)^k < \epsilon$$

for c sufficiently large, with $\text{Arg } c = \alpha$. Let D_2 be such that $|b_t - c_t| < D_2/2$ for all sufficiently large t . Let $D = D_1 + D_2$. Then

$$\begin{aligned} (114) \quad f(x, t) &= F(x, c_t) = \sum_{k=0}^{\infty} F_k(c_t)(x - c_t)^k \\ &= \sum_{n=0}^{\infty} (x - b_t)^n \sum_{k=n}^{\infty} F_k(c_t) \binom{k}{n} (b_t - c_t)^{k-n} \\ &= \sum_{n=0}^{\infty} (x - b_t)^n f_n(t). \end{aligned}$$

Hence

$$\begin{aligned} (115) \quad \sum_{n=1}^{\infty} |f_n(t)| (|b_t| - D)^n &\leq \sum_{k=1}^{\infty} |F_k(c_t)| (|b_t| - D + |b_t - c_t|)^k \\ &\leq \sum_{k=1}^{\infty} |F_k(c_t)| (|c_t| - D_1)^k < \epsilon. \end{aligned}$$

Also, since $|f_0(t)| \leq |F(c_t, c_t)| + \sum_{k=0}^{\infty} |F_k(c_t)| (D_2/2)^k$, $f_0(t)$ is bounded. Hence $f(x, t)$ is almost constant on S .

LEMMA XV. (1) Let $f(x, t)$ and $g(x, t)$ be almost constant on a $Q(\alpha)$ -sequence S . Then $f(x, t) + g(x, t)$ and $f(x, t)g(x, t)$ are almost constant on S .

(2) Let $F(x, b)$, $G(x, b)$ be almost constant in direction α . Then $F(x, b) + G(x, b)$ and $F(x, b)G(x, b)$ are almost constant in direction α .

Proof. Obvious.

LEMMA XVI. Let $\phi(x)$ be analytic at ∞ , and in the domain $|x| \geq C$. Then there is a positive M such that $|\phi^{(k)}(b)| \leq M k! / (|b| - C)^{k+1}$ if $k > 0$ and $|b| > C$.

Proof. Let $|\phi(x)| \leq M_1$ in the set $|x| \geq C$. Then, if $\phi(x) = \sum_{j=0}^{\infty} a_j x^{-j}$, we have $|a_j| \leq M_1 C^j$. Let $|b| > C$. Let $u = x - b$. Then

$$(116) \quad \phi(x) = \sum_{j=0}^{\infty} a_j (u + b)^{-j} = \sum_{k=0}^{\infty} u^k \sum_{j=0}^{\infty} a_j \binom{-j}{k} b^{-(j+k)}.$$

Thus

$$(117) \quad |\phi^{(k)}(b)| \leq k! \sum_{j=0}^{\infty} |a_j| \left| \binom{-j}{k} \right| |b|^{-(j+k)} \\ \leq k! \sum_{j=0}^{\infty} M_1 C^j \left| \binom{-j}{k} \right| |b|^{-(j+k)}.$$

Let $H(x) = M_1 - (M_1 C)/(x + C) = M_1 x/(x + C)$. Let $v = x + |b|$. Then

$$(118) \quad H(x) = \sum_{k=0}^{\infty} v^k \sum_{j=0}^{\infty} M_1 C^j |b|^{-(j+k)} \left| \binom{-j}{k} \right|.$$

Hence $|\phi^{(k)}(b)| \leq H^{(k)}(-|b|) = k! M_1 C (|b| - C)^{-(k+1)}$, which establishes the formula with $M = M_1 C$.

LEMMA XVII. Every function analytic at ∞ is almost constant in direction α , for every α .

Proof. This follows immediately from Lemma XVI.

LEMMA XVIII. Let (q_t, b_t) , $(t = 1, 2, \dots)$, be a $Q(\alpha)$ -sequence. Let C be a positive number. Let \mathcal{S}_t be the point-set $|x - b_t| < \mathcal{R}(b_t e^{-i\alpha}) - C$. Then if \mathcal{F} is any closed bounded set such that $\mathcal{R}(x e^{-i\alpha}) > C$ at every x of \mathcal{F} , there is a positive T such that \mathcal{F} is included in \mathcal{S}_t whenever $t > T$.

Proof. Let $\sigma = e^{-i\alpha}$. Then

$$(119) \quad |x - b|^2 = |x\sigma - b\sigma|^2 = [\mathcal{R}(\sigma b) - C]^2 + \{\mathcal{R}(\sigma b)\} \cdot \\ \{[2C - 2\mathcal{R}(\sigma x)] + [\tan(\text{Arg } b - \alpha)\mathfrak{I}(b\sigma) \\ - 2\mathfrak{I}(x\sigma)\tan(\text{Arg } b - \alpha) + (|x|^2 - C^2)/\mathcal{R}(\sigma b)]\}.$$

Since $\lim_{t \rightarrow \infty} \mathcal{R}(b\sigma) = \lim_{t \rightarrow \infty} |b| \cos(\text{Arg } b - \alpha) = +\infty$, and $\tan(\text{Arg } b - \alpha)$

$= O(|b|^{-1})$, while $\Im(b\sigma) = |b| \sin(\text{Arg } b - \alpha) = O(1)$, it follows from equation (119) that there is a positive T such that if $t > T$, then $|x - b|^2 < (\Re(\sigma b) - C)^2$ for all x in \mathcal{F} .

LEMMA XIX. *If $f(x)$ is almost constant in direction α , then $f(x)$ tends to a finite limit as the real part of $xe^{-i\alpha}$ tends to positive infinity.*

Proof. Let ϵ be any positive number. Let D be such that if $|b| > D$, and $\text{Arg } b = \alpha$, and $f(x) = \sum_{k=0}^{\infty} f_k(b)(x - b)^k$, then

$$\sum_{k=1}^{\infty} |f_k(b)| (|b| - D)^k < \epsilon.$$

Let x_1, x_2 be any two complex numbers such that $\Re(x_1 e^{-i\alpha}) > D$ and $\Re(x_2 e^{-i\alpha}) > D$.

Let b be such that $|b| > D$ and $\text{Arg } b = \alpha$. Now

$$\begin{aligned} |f(x_1) - f(x_2)| &\leq |f(x_1) - f(b)| + |f(b) - f(x_2)| \\ &= \left| \sum_{k=1}^{\infty} f_k(b)(x_1 - b)^k \right| + \left| \sum_{k=1}^{\infty} f_k(b)(b - x_2)^k \right| \leq \sum_{k=1}^{\infty} |f_k(b)| |x_1 - b|^k \\ &\quad + \sum_{k=1}^{\infty} |f_k(b)| |x_2 - b|^k < 2\epsilon \end{aligned}$$

if b is sufficiently large, since $|x_1 - b| < |b| - D$, and $|x_2 - b| < |b| - D$, if b is sufficiently large (by Lemma XVIII).

LEMMA XX. *Let $F(x)$ be bounded and analytic in a right half-plane, $\Re(x) > D > 0$, and on the boundary of that half-plane. Let b be a positive number, and let $q = 1 - 1/b$. Let ω be a fixed complex number of positive real part. Let $V = b(1 - q^\omega)$, where $q^\omega = e^{\omega \log q}$.*

Let D_1 be any positive number greater than D . Then there exists a positive number M (independent of b and x), such that

$$|F(q^\omega x + V) - F(x + \omega)| < (M|x|)/(b\Re(x))$$

for all sufficiently large b and all x in the circle $|x - b| \leq b - D_1$.

Proof. Let x be any point of the circle $|x - b| \leq b - D_1$. Then $x = b + \rho e^{i\theta}$, where $0 \leq \rho \leq b - D_1$, and $-\pi < \theta \leq \pi$. Hence

$$\begin{aligned} |\tan \text{Arg } x| &= |\rho \sin \theta (b + \rho \cos \theta)^{-1}| \leq (\rho)^{\frac{1}{2}} (2D_1)^{\frac{1}{2}} \\ &\leq (b)^{\frac{1}{2}} (2D_1)^{-\frac{1}{2}}. \end{aligned}$$

Now $\Re(q^\omega x + V) = \Re(q^\omega x) + \Re(V)$. Since V tends to the limit ω as b becomes infinite, it follows that $\Re(V) > 0$ if b is large. Hence $\Re(q^\omega x + V) > \Re(q^\omega x)$ if b is large. Let $q^\omega = Re^{i\psi}$, with $R > 0$ and $-\pi < \psi \leq \pi$. Then limit $\psi/\text{Log } R = \tan(\text{Arg } \omega)$, by Lemma III. Now $\text{Log } R = O(b^{-1})$, and $\psi = O(b^{-1})$. Hence $\sin \psi = O(b^{-1})$, and $1 - \cos \psi = O(b^{-2})$. We have

$$(120) \quad \begin{aligned} \Re(q^\omega x) &= R |x| \cos(\psi + \text{Arg } x) \\ &= R |x| [\cos \psi \cos(\text{Arg } x)] (1 - \tan \psi \tan \text{Arg } x). \end{aligned}$$

Now $\tan \psi \tan \text{Arg } x \leq M_1(b^{-1})(b)^{1/2}$, for some positive M_1 (independent of b and x). Let A be any positive number less than 1. Then it follows from (120) that $\Re(q^\omega x) > A \Re(x)$ if b is sufficiently large. Hence, if b is sufficiently large, and x lies in the circle $|x - b| < b - D_1$ then $q^\omega x + V$ lies in the half-plane $\Re(x) > D$. Now

$$(121) \quad \begin{aligned} F(q^\omega x + V) - F(x + \omega) \\ = (2\pi i)^{-1} \int_C \frac{F(\xi)[(q^\omega x + V) - (x + \omega)]}{(\xi - (q^\omega x + V))(\xi - (x + \omega))} d\xi \end{aligned}$$

where C is any circle lying in the half-plane $\Re(x) \geq D$ and containing $q^\omega x + V$ and $x + \omega$ in its interior. We take C to be the circle with center $x + \omega$ and radius $r = \Re(x + \omega) - D$. Then $q^\omega x + V$ will lie in C provided $|(x + \omega) - (q^\omega x + V)| < r$. There exists a positive number A_1 such that $|(x + \omega) - (q^\omega x + V)| < A_1 |x|/b$, for all x , since the relation $q = 1 - 1/b$ implies that $\omega - V = O(b^{-1})$. Hence to show that $q^\omega x + V$ lies in C it suffices to show that $A_1 |x|/b < \Re(x + \omega) - D$. Since there is a positive number A_2 such that $\Re(x + \omega) - D > A_2 \Re(x)$ when $\Re(x) \geq D_1$, it suffices to show that $A_1 |x|/b < A_2 \Re(x)$, or that $b \cos(\text{Arg } x) > A_1/A_2$. Since $\cos(\text{Arg } x) = (1 + \tan^2 \text{Arg } x)^{-1/2} > (1 + b/2D_1)^{-1/2}$, when $|x - b| < b - D_1$, we conclude that if b is sufficiently large, and $|x - b| < b - D_1$, then $b \cos(\text{Arg } x) > A_1/A_2$. It follows that if b is sufficiently large, and $|x - b| < b - D_1$, then $q^\omega x + V$ is in C .

By the same argument, if b is sufficiently large, then $|(x + \omega) - (q^\omega x + V)| < r/2$ for all x in the set $|x - b| < b - D_1$. Then if $|F(x)| \leq M_1$ in the half-plane $\Re(x) \geq D$, we have

$$(122) \quad \begin{aligned} |F(q^\omega x + V) - F(x + \omega)| \\ < (2\pi)^{-1} \int_0^{2\pi} \frac{M_1(A_1 |x|/b)r d\theta}{(r/2)A_2 \Re(x)} = (M |x|)/(b \Re(x)). \end{aligned}$$

LEMMA XXI. Let $G(x, b)$ be a function of x and b such that, for some

positive D_1 , $G(x, b)$ is analytic in the circle $|x - b| < b - D_1$ for all sufficiently large positive b , and in that circle satisfies an inequality $|G(x, b)| < M|x|/b\Re(x)$. Then $G(x, b)$ is almost constant in direction 0, and moreover, for every $D_2 > D_1$, there is a positive number M_2 such that if $G(x, b) = \sum_{\lambda=0}^{\infty} G_{\lambda}(x-b)^{\lambda}$, then $|\sum_{\lambda=0}^{\infty} G_{\lambda}|u^{\lambda}| < M_2 b^{-1/4}$ if $|u| < b - D_2$.

Proof. Let $G(x, b) = \sum_{\lambda=0}^{\infty} G_{\lambda}u^{\lambda}$, where $u = x - b$. Let D_3 be such that $D_2 > D_3 > D_1$. Let $\rho = b - D_3$. If $|u| < b - D_2$, then

$$\begin{aligned} (123) \quad & \left| \sum_{\lambda=0}^{\infty} G_{\lambda}|u^{\lambda}|^2 \right| = \left| \sum_{\lambda=0}^{\infty} G_{\lambda}|\rho^{\lambda}(u/\rho)^{\lambda}|^2 \right| \\ & \leq \left(\sum_{\lambda=0}^{\infty} |G_{\lambda}|^2 \rho^{2\lambda} \right) \left(\sum_{\lambda=0}^{\infty} |u/\rho|^{2\lambda} \right) \\ & = (1 - |u/\rho|^2)^{-1} (2\pi)^{-1} \int_0^{2\pi} |G(b + \rho e^{i\theta})|^2 d\theta \\ & \leq (1 - |u/\rho|^2)^{-1} (2\pi)^{-1} \int_0^{2\pi} \frac{M^2 |b + \rho e^{i\theta}|^2}{b^2 \Re^2(b + \rho e^{i\theta})} d\theta \\ & = (1 - |u/\rho|^2)^{-1} M^2 b^{-1} (b^2 - \rho^2)^{1/2} < M^2 (D_2 - D_3)^{-1} D_3^{-1/2} b^{-1/2}. \end{aligned}$$

LEMMA XXII. Let $f(x, y_1, \dots, y_n)$ be a polynomial in the y_k , with coefficients functions of x analytic and bounded in a right half-plane. Let $Y(x)$ be a function of x analytic and bounded in a right half-plane. Let $\omega_1, \dots, \omega_n$ be complex numbers, with $\omega_1 = 0$ and $\Re(\omega_k) > 0$ when $k > 1$. Let b be a positive number, let $q = 1 - 1/b$, and let $V_k = b(1 - q^{\omega_k})$, ($k = 1, \dots, n$). Let

$$\begin{aligned} E(x, b) &= f(x, Y(x + \omega_1), \dots, Y(x + \omega_n)) \\ &= f(x, Y(q^{\omega_1}x + V_1), \dots, Y(q^{\omega_n}x + V_n)). \end{aligned}$$

Then $E(x, b)$ is almost constant in direction 0, and moreover if $E(x, b) = \sum_{\lambda=0}^{\infty} E_{\lambda}(x-b)^{\lambda}$, then there exists a positive D_3 and a positive M_3 , both independent of b , such that

$$(124) \quad \left| \sum_{\lambda=0}^{\infty} E_{\lambda}|u^{\lambda}| \right| < M_3 b^{-1/4}$$

if $|u| < b - D_3$ and b is sufficiently large.

Proof. Directly from its definition $E(x, b)$ is the sum of finitely many terms of the form

$$(125) \quad a(x) \left\{ \prod_{k=1}^n [Y(x + \omega_k)]^{i_k} - \prod_{k=1}^n [Y(q^{\omega_k}x + V_k)]^{i_k} \right\}$$

where $a(x)$ is analytic and bounded in a right half-plane, and i_1, \dots, i_n are non-negative integers such that $i_1 + \dots + i_n \geq 1$. Such a term may be written in the form

$$(126) \quad a(x) \sum_{s=1}^M h_s(x) [Y(x + \omega_{k(s)}) - Y(q^{\omega_{k(s)}}x + V_{k(s)})],$$

where $M = i_1 + \dots + i_n$, and $k(s)$ is for each s an integer between 1 and n , and $h_s(x)$ is a power product in the $Y(x + \omega)$ and the $Y(q^\omega x + V)$. Thus, since $a(x)$ and each $h_s(x)$ is bounded, it follows from Lemma XX that $E(x, b)$ satisfies an inequality $|E(x, b)| < M_4 |x| (b\Re(x))^{-1}$ in some circle $|x - b| < b - D_4$. Hence Lemma XXI is applicable, to establish the present theorem.

Section E. Special lemmas.

LEMMA XXIII. Let b_1, b_2, \dots be a sequence of complex numbers tending to ∞ . Let $\lambda_1, \lambda_2, \dots$ be a sequence of non-negative numbers, such that $\lim_{t \rightarrow \infty} (\lambda_t/b_t) = \infty$. Let c_1, c_2, \dots be any sequence of complex numbers. Let $f(x, t) = c_t(x - b_t)^{\lambda_t}$, in a simply connected region \mathcal{B} containing none of the points b_t . (Any determination of $(x - b_t)^{\lambda_t}$ being used.)

Then if $\lim_{t \rightarrow \infty} f(x, t)$ exists, uniformly in an open subset \mathcal{B}_0 of \mathcal{B} , the limit function is identically zero in \mathcal{B}_0 .

Proof. Evidently $f'(x, t)/f(x, t) = \lambda_t/(x - b_t)$, where $f'(x, t)$ is the derivative of $f(x, t)$ with respect to x . Hence $f'(x, t)/f(x, t)$ becomes infinite as t becomes infinite. This precludes the possibility that $f(x, t)$ tend to a non-zero finite limit at a point of \mathcal{B}_0 .

LEMMA XXIV. Let $\alpha_k = \text{Arg}(\text{Log } 2 + 2\pi ki)$, ($k = 0, \pm 1, \pm 2, \dots$). Let $h(x)$ be a solution of the equation

$$(127) \quad h(x) - 2h(x+1) = 0,$$

analytic in a half-plane $\Re(xe^{-i\alpha}) > D > 0$, and satisfying there a condition $|h(x)| \leq M|x|^N$ for some positive M, N .

Then $h(x)$ is bounded in $\Re(xe^{-i\alpha}) > D$, and if α is different from every α_k , then $h(x) \equiv 0$, while if $\alpha = \alpha_k$, then $h(x) = ce^{\sigma_k x}$ for some constant c , and for $\sigma_k = \text{Log}(1/2) + 2\pi ki$.

Proof. Case 1. $\cos \alpha \neq 0$. In this case equation (127) provides an analytic continuation of $h(x)$ throughout the entire plane. Let $f(x) = 2^x h(x)$. Then $f(x)$ is an entire function of period 1. Hence $f(x) = \sum_{n=-\infty}^{\infty} a_n e^{2\pi i n x}$. Let $z = e^{2\pi i x}$, and let $G(z) = f(x) = \sum_{n=-\infty}^{\infty} a_n z^n$. Then $G(z)$ is analytic for all finite z except $z = 0$. Suppose there is a j such that $a_j \neq 0$. Let $H(z) = z^{-j} G(z) = \sum_{n=-\infty}^{\infty} a_n z^{n-j}$. Let $r = |z|$, and let r be large. Then there is a positive C such that if z is chosen suitably, with $|z|$ large, then $|z^{-j} G(z)| \geq C$. That is, $|G(z)| \geq C r^j$. In other words, if x is chosen suitably, with $|e^{2\pi i x}| = r$, and r large, then $|f(x)| \geq C r^j$.

On the boundary of the half-plane $\Re(xe^{-i\alpha}) > D$ we have $\Re(x) = D \sec \alpha - \Im(x) \tan \alpha$. Let $\Im(x) = -(\log r)/(2\pi)$. Then $|e^{2\pi i x}| = r$. Hence for some x with $\Im(x) = -(\log r)/(2\pi)$, we have $|f(x)| \geq C r^j$. Because of the periodicity of $f(x)$, we may take this x to have real part equal to $D \sec \alpha + ((\log r)(\tan \alpha)/(2\pi)) + \theta$, where $0 < \theta \leq 1$, if $\cos \alpha > 0$, and $-1 \leq \theta < 0$ if $\cos \alpha < 0$. Then x lies in the half-plane $\Re(xe^{-i\alpha}) > D$. Hence

$$(128) \quad |h(x)| \leq M |x|^N \leq M_1 (\log r)^N, \text{ for some positive } M_1.$$

But

$$(129) \quad |h(x)| = |2^{-x}| |f(x)| \geq \exp(-\Re(x)(\log 2)) C r^j \\ \geq C_1 \exp[(\log r)(j - (\log 2)(\tan \alpha)/(2\pi))].$$

It follows that $j \leq (\log 2)(\tan \alpha)/(2\pi)$.

By a similar argument, using r sufficiently small, one shows that if $a_j \neq 0$, then $j \geq (\log 2)(\tan \alpha)/(2\pi)$.

Thus, if $a_j \neq 0$, then $j = (\log 2)(\tan \alpha)/(2\pi)$. Hence $h(x) \equiv 0$, unless there is an integer k such that $\tan \alpha = (2\pi k)/(\log 2)$; if $\tan \alpha = (2\pi k)/(\log 2)$, then $h(x) \equiv c 2^{-x} e^{2\pi i k x} = c e^{\sigma_k x}$, for some constant c .

Now if $\tan \alpha = (2\pi k)/(\log 2)$, either $\alpha = \alpha_k$, or $\alpha \equiv \alpha_k + \pi \pmod{2\pi}$. For the second possibility we must have $c = 0$, because if $c \neq 0$, then $c e^{\sigma_k x}$ is evidently not majorized by $M |x|^N$ in the half-plane $\Re(xe^{-i\beta}) > D$ when $\beta \equiv \alpha_k + \pi \pmod{2\pi}$. Thus the lemma follows, in Case 1.

Case 2. $\cos \alpha = 0$. If $h(x_0) = A \neq 0$, then $h(x_0 - n) = 2^n A$, which contradicts the hypothesis $|h(x_0 - n)| \leq M |x_0 - n|^N$.

ON TOPLER'S WAVE ANALYSIS.*

By AUREL WINTNER.

In a quite forgotten paper [2], appearing in a periodical which was not generally accessible even at the time of its publication (1872), the physicist Töpler¹ has dealt with a generalization of harmonic analysis² (on a finite time-range). The generalization in question consists in replacing the overtones of $\sin t$ and/or $\cos t$ by the sequence $\phi(t), \phi(2t), \phi(3t), \dots$, where $\phi(t)$ is an "arbitrary" periodic function.

As implied by context *and* date, Töpler's considerations on such a basic sequence are purely formal in nature, even though some of his formulae are quite ingenious. From what was available at that time, he could have, perhaps, derived more solid foundations from a suggestion made to him by Boltzmann, which he mentions but does not follow further. Boltzmann's suggestion is that $\phi(t)$ should be expanded into its ordinary Fourier series, in which t should then be replaced by $2t, 3t, \dots$; thus obtaining a (non-recursive) system of an infinity of linear equations (to be solved), connecting the functions $\phi(nt)$ with the functions $\sin mt$ and/or $\cos mt$.

This program can be carried out today. In its full generality, which leads to very simple results but without which the theory would become cumbersome, it could not have been carried out at that time. In fact, recourse will have to be had not only to Lebesgue's definition of an integral (without which there is no Fischer-Riesz theorem) but also to Hilbert's theory of bounded matrices. What will be needed in the latter regard is the theory of bounded "*D*-matrices," introduced by Toeplitz [3]; cf. also the Appendix of reference 2) in [4].

* Received December 10, 1946.

¹ To the mathematician, Töpler's name is familiar in connection with the quadratic extremal property of Fourier constants; a connection in which he is mentioned, for instance, on p. 639 of vol. 2 of Hobson's *Theory of Functions* (2nd ed.).

² In [5], I failed to refer to Töpler's paper, which I knew well twenty years earlier but which I had not seen in the meantime. I must have forgotten it, even though I clearly was under its influence (in particular in connection with Boltzmann's suggestion). I realized this only recently, while turning over Burkhardt's report [1].

In the literature known to me, Burkhardt's abstract (*loc. cit.*, pp. 905-906) is the only passage referring to Töpler's paper [2].

As pointed out in [4], p. 147, what is actually accomplished by Toeplitz's Dirichlet matrices is a representation of the Eratosthenian sieve process. Hence, the situation agrees with Töpler's prediction of arithmetical implications ([2], p. 98). On the other hand, it cannot be surprising that certain assertions made, anno 1872, by a physicist interested in wave analysis should prove to be inaccurate (wrong is, for instance, the criterion claimed in [2], p. 71, concerning the sufficiency of absolute convergence).

Actually, the whole problem becomes well-defined only if it is specified whether true *series* (with unique coefficients) or just approximability by *sequences* (that is to say, the "separation" property) is required, and there result two additional questions if the (L^2) -metric is replaced by the l. u. b.-metric of uniform convergence (the corresponding questions concerning mere convergence, or convergence almost everywhere, are of course at least as intricate, hence "special," as in the particular case of ordinary Fourier series). It would be hard to say or, rather, it would be unhistorical to ask, which of these possibilities Töpler and Boltzmann actually had in mind.

With reference to a sequence of constants ϕ_1, ϕ_2, \dots , let $D(\Phi)$ denote the infinite matrix in which the indices of the rows and the columns range through the positive integers and the n -th column is formed by the sequence

$$(1) \quad (0, \dots, 0, \phi_1, 0, \dots, 0, \phi_2, 0, \dots, 0, \phi_3, \dots),$$

where each block of 0's is of length $n-1$ (it being understood that these blocks are considered as missing when their length is 0, i. e., in the first column). Since $(\phi_1, \phi_2, \phi_3, \dots)$ is the first, $(0, \phi_1, 0, \phi_2, 0, \dots)$ the second, $(0, 0, \phi_1, 0, 0, \phi_2, \dots)$ the third, \dots column of $D(\Phi)$, every element of $D(\Phi)$ situated above the principal diagonal is 0 and every element in the principal diagonal is ϕ_1 . The n -th component of the vector, say (f_1, f_2, \dots) , into which a vector, say (c_1, c_2, \dots) , is transformed by $D(\Phi)$ is

$$(2) \quad f_n = \sum_{d|n} \phi_d c_{n/d},$$

where d runs through all divisors (≥ 1) of n .

In Toeplitz's paper [3], the matrix of his D -form is chosen to be the transposed matrix of the matrix of the linear substitution (2); so that (2) becomes replaced by

$$(2 \text{ bis}) \quad f'_n = \sum_{m=1}^{\infty} \phi_n c_{nm}$$

(nm is the product of n and m , and not a double subscript). In contrast to

(2), the transposed substitution, (2 bis), is defined for every (c_1, c_2, \dots) satisfying

$$(3) \quad \sum_{n=1}^{\infty} |c_n|^2 < \infty$$

if and only if

$$(4) \quad \sum_{m=1}^{\infty} |\phi_m|^2 < \infty.$$

If (2 bis) is used instead of (2), the Eratosthenian connection becomes obscured. However, since the properties of a matrix with regard to Hilbert's space are identical with those of the transposed matrix, Toeplitz's theory of D -forms can be formulated in terms of (2 bis) as well as in terms of (2).

When expressed in terms of the latter, the Eratosthenian, algorithm, Toeplitz's principal results in [3] are that

(i) the matrix $D(\Phi)$ is bounded (in Hilbert's sense) if and only if the ordinary Dirichlet series

$$(5) \quad \Phi(s) = \sum_{n=1}^{\infty} \phi_n/n^s$$

is convergent in the half-plane $\sigma > 0$ and satisfies the condition

$$(6) \quad |\Phi(s)| < \text{const.}, \text{ where } \sigma > 0,$$

$(s = \sigma + it)$; that

(ii) if the assumptions of (i) are satisfied and, in addition,

$$(7) \quad |1/\Phi(s)| < \text{Const.}, \text{ where } \sigma > 0,$$

then $D(\Phi)$ has a (unique) bounded reciprocal matrix, $D^{-1}(\Phi)$, which is precisely $D(\Psi)$, where $\Psi(s) = 1/\Phi(s)$; a fact which, as shown in [4], can be refined by saying that

(ii bis) if the assumptions of (i) are satisfied, then the spectrum of $D(\Phi)$ (meant in the sense in which I defined the spectrum of any, not necessarily Hermitian, bounded matrix) is identical with the closure of the values attained by the function (5) in the half-plane $\sigma > 0$.

In addition, use will be made of the following pair of facts, in which A^* denotes \bar{A}' , if \bar{A} is the complex conjugate, and A' the transposed matrix, of a matrix A (so that, in particular, $z^* = \bar{z} = x - iy$, if $z = x + iy$ is a number):

(α) If $A^* = A$, where $A = (a_{nm})$, then the Hermitian form

$$\sum_{n=1}^{\infty} \sum_{m=1}^{\infty} a_{nm} c_n c_m^*,$$

when thought of as a double series with the restriction that, in its partial sums, the index pair (n, m) occurs whenever (m, n) occurs, is convergent at every point (c_1, c_2, \dots) of Hilbert's space if and only if A is bounded.

(β) An arbitrary matrix $A = (a_{nm})$, where $n, m = 1, 2, \dots$, is bounded if and only if either the product AA^* or the product A^*A (exists and) is bounded.

Both (α) and (β) are classical (Toeplitz, Hellinger-Toeplitz); cf. pp. 121-132 of my *Spektraltheorie*, 1929).

From (β) and (i), it is easy to deduce the following criterion (cf. [5], pp. 575-576), the proof of which will be detailed for the sake of completeness:

(I) Let $\phi(t)$ be a function of class (L^2) on the t -interval $(0, \pi)$, i. e., let (4) be satisfied by the constants ϕ_1, ϕ_2, \dots which result by placing

$$(8) \quad \phi(t) \sim \sum_{n=1}^{\infty} \phi_n \sin nt$$

on $(0, \pi)$. Then the infinite Hermitian matrix

$$(9) \quad \left(\int_0^{\pi} \phi(nt) \phi^*(mt) dt \right), \quad (n, m = 1, 2, \dots),$$

where $\phi(t)$ is meant to be defined for $-\infty < t < \infty$ by (8), is a bounded matrix if and only if the Dirichlet series (5) is convergent in the half-plane $\sigma > 0$ and satisfies (6).

Although

$$(10) \quad \int_0^{\pi} \phi(t) dt = 0$$

is not assumed, the summation index $n = 0$ does not occur in (8). It does occur in the corresponding cosine series,

$$(11) \quad \phi(t) \sim \sum_{n=0}^{\infty} \phi_n \cos nt,$$

of $\phi(t)$ on $(0, \pi)$. However, it does not occur in, or does not have any

influence on, the boundedness of either (5) or (9), and it will be clear from the proof of (I) that

(I bis) (8) can be replaced by (11) in (I), provided that (10) is satisfied (i. e., if $\phi_0 = 0$).

The case of

$$(12) \quad \phi(t) \sim \frac{1}{2}a_0 + \sum_{n=1}^{\infty} (a_n \cos nt + b_n \sin nt),$$

where (instead of $0 < t < \pi$)

$$(13) \quad 0 < t < 2\pi,$$

can, of course, also be treated.

According to (8),

$$\phi(nt) \sim \sum_{k=1}^{\infty} \phi_k \sin knt$$

and (z^* denoting the complex conjugate of z)

$$\phi^*(mt) \sim \sum_{k=1}^{\infty} \phi_k^* \sin kmt.$$

Hence, an application of the ("bilinear" form of the "quadratic") Parseval relation shows that the integral occurring in (9) is a fixed multiple of the sum

$$(14) \quad \sum_{k=1}^{\infty} \phi_{km/(n,m)} \phi_{kn/(n,m)}^*,$$

where (n, m) denotes the greatest common divisor of n and m (and divides, therefore, the numerators $n, m, 2n, 2m, \dots, kn, km, \dots$ of the indices occurring in this bilinear form of ϕ_i and ϕ_j^*). On the other hand, since $D(\Phi)$ has been defined as the matrix the n -th column of which is the sequence (1), the sum (14) clearly is the complex conjugate matrix element at intersection of the n -th row and the m -th column of the product $D^*(\Phi)D(\Phi)$, the last asterisk being the one defined before (α). It follows therefore from (β) that the matrix (9) is bounded if and only if the matrix $D(\Phi)$ is. Consequently, (I) follows from (i).

It will now be shown that, if (α) is used instead of (β), there results from (I) the following theorem:

(II) Let $\phi(t)$ be a function defined for $-\infty < t < \infty$ by (8), where ϕ_1, ϕ_2, \dots is a sequence of constants satisfying (4). Then the partial sums,

$$(16) \quad f^k(t) = \sum_{n=1}^k c_n \phi(nt),$$

of the infinite series

$$(17) \quad \sum_{n=1}^{\infty} c_n \phi(nt)$$

belonging to any sequence (c_1, c_2, \dots) satisfying (3) tend in the mean (L^2) to a function

$$(17 \text{ bis}) \quad f(t) = f(t; c_1, c_2, \dots),$$

(of class (L^2) on $(0, \pi)$) if and only if the Dirichlet series (5) is convergent in the half-plane $\sigma > 0$ and satisfies (6).

In other words, there belongs to every or not to every point (c_1, c_2, \dots) of Hilbert's space a function $f(t)$ satisfying

$$(18) \quad \int_0^{\pi} |f(t) - f^k(t)|^2 dt \rightarrow 0, \quad k \rightarrow \infty,$$

according as the coefficients, ϕ_n , of (8) do or do not satisfy the requirements of (I).

Corresponding to the cosine analogue, (I bis), of (I), there is an analogue of (II):

(II bis) (8) can be replaced by (11) in (II), provided that (10) is satisfied (i. e., if $\phi_0 = 0$).

First, if c_1, c_2, \dots is a fixed sequence of constants which need not satisfy (3), it follows from the completeness of the (L^2)-space that the existence of an $f(x)$ satisfying (18) is equivalent to

$$\int_0^{\pi} |f^j(t) - f^k(t)|^2 dt \rightarrow 0, \quad j, k \rightarrow \infty.$$

In view of (16), this can be written in the form

$$\sum_{n=k+1}^j \sum_{m=k+1}^j c_n c_m^* \int_0^{\pi} \phi(nt) \phi^*(mt) dt \rightarrow 0, \quad j, k \rightarrow \infty,$$

where $k \leq j$. But the expression on the left of this limit relation is identical with the (k, j) -th remainder term of the *restricted* double series mentioned in (α), if $A = A^*$ in (α) is identified with the matrix (9). Hence, (II) follows from (α) and (I).

By using (ii) instead of (i), it will now be easy to deduce the following solution of the Töpler-Boltzmann problem in the (L^2)-case:

(III) Let ϕ_1, ϕ_2, \dots be a sequence of constants having the property that the Dirichlet series (5) converges in the half-plane $\sigma > 0$ and satisfies both (6) and (7). Then (8) defines a function which is of class (L^2) and has the property that the relation

$$(19) \quad f(t) \sim \sum_{n=1}^{\infty} c_n \phi(nt),$$

when interpreted as an expansion in the mean (L^2) [cf. (18), (16)], establishes a one-to-one correspondence between all sequences (c_1, c_2, \dots) satisfying (3) and all functions $f(t)$ of class (L^2) on $(0, \pi)$.

The point is that the sequence

$$(20) \quad \phi(t), \phi(2t), \dots, \phi(nt), \dots$$

does not (in general) consist of orthogonal functions; so that the coefficients of (19) cannot be obtained in Fourier's fashion, and their characterization in terms of the minimum property, referred to in the first footnote, is not available.

What is claimed in (III) but not in (II) contains two corollaries neither of which is obvious; namely, that

(III') under the assumptions of (III), the sequence (20) is a basis of the functions $f(t)$ of class (L^2) on $(0, \pi)$,

and that

(III'') under the assumptions of (III),

$$(21) \quad 0 \sim \sum_{n=1}^{\infty} c_n \phi_n(nt) \text{ only when } \sum_{n=1}^{\infty} |c_n|^2 = 0 \quad (\text{if } \sum_{n=1}^{\infty} |c_n|^2 < \infty);$$

in particular, the functions (20) are linearly independent.

(The latter statement is weaker than (21), since all that it claims is that, when k is finite,

$$(21^*) \quad \sum_{n=1}^k |c_n|^2 = 0, \text{ if } \sum_{n=1}^k c_n \phi(nt) = 0$$

holds almost everywhere.)

The basis assertion of (III') means that, if an $f(t)$ of class (L^2) and an $\epsilon > 0$ are given, there exist a k and k constants C_n satisfying

$$\int_0^{\pi} |f(t) - \sum_{n=1}^k C_n \phi(nt)|^2 dt < \epsilon.$$

This is much less than (19), since the minimum property of the coefficients of (19) is not available. It should be noted in this regard that the assertion of the last formula line has been proved in [5], p. 566 and pp. 572-573, for the case

$$\phi(t) \sim \sum_{n=1}^{\infty} n^{-\lambda} \sin nt = \phi(t)$$

if $\lambda > \frac{1}{2}$, whereas the assertions of (III) for (19) are satisfied only if $\lambda > 1$.

The proof of (III) will involve (II). If (II bis) is used instead of (II), what results is the following variant of (III):

(III bis) (8) can be replaced by (11) in (III), provided that (19) is replaced by

$$(19 \text{ bis}) \quad f(t) \sim c_0 + \sum_{n=1}^{\infty} c_n \phi(nt),$$

and (c_1, c_2, \dots) by (c_0, c_1, c_2, \dots) . In particular, (20) in (III') can then be replaced by

$$(20 \text{ bis}) \quad 1, \phi(t), \phi(2t), \dots, \phi(nt), \dots$$

According to (i) and (ii), the assumptions of (III) mean that $D(\Phi)$ is a bounded matrix and has a (unique) bounded reciprocal matrix $D^{-1}(\Phi)$. But the boundedness of a matrix means that the matrix transforms every point of Hilbert's space into a point of Hilbert's space (Toeplitz). Hence, if (f_1, f_2, \dots) is any vector satisfying

$$(22) \quad \sum_{n=1}^{\infty} |f_n|^2 < \infty,$$

it is transformed by $D^{-1}(\Phi)$ into a vector,

$$(23) \quad (c_1, c_2, \dots) = D^{-1}(\Phi)(f_1, f_2, \dots),$$

satisfying (3). Furthermore, this mapping of the Hilbert space (f) on the Hilbert space (c) has a unique inverse on the latter space, since (23) has a (unique) bounded inverse,

$$(24) \quad (f_1, f_2, \dots) = D(\Phi)(c_1, c_2, \dots).$$

Since $\sin t, \sin 2t, \dots$ is a complete orthogonal system on $(0, \pi)$, the Fischer-Riesz assignment,

$$(25) \quad f(t) \sim \sum_{n=1}^{\infty} f_n \sin nt,$$

establishes a one-to-one correspondence between the sequences (f_1, f_2, \dots) satisfying (22) and the functions $f(t)$ of class (L^2) on $(0, \pi)$. Consequently, (24) and (2) imply that the n -th Fourier sine constant, f_n , of every function, $f(t)$, of class (L^2) on $(0, \pi)$ can be represented in the form

$$(26) \quad f_n = \sum_{d|n} c_d \phi_{n/d},$$

where c_1, c_2, \dots is a sequence satisfying (3). On the other hand, (II) states that any such sequence defines a function to which the partial sums, (16), of (17) tend in the mean (L^2) on $(0, \pi)$. Let $g(t) = g(t; c_1, c_2, \dots)$ denote this limit function of class (L^2) . Then, since

$$\int_0^\pi |g(t) - f^k(t)|^2 dt \rightarrow 0 \text{ as } k \rightarrow \infty,$$

holds for the functions (16) and since, if $g(t) \sim \sum_{n=1}^\infty g_n \sin nt$,

$$\int_0^\pi |g(t) - \sum_{m=1}^j g_m \sin mt|^2 dt \rightarrow 0 \text{ as } j \rightarrow \infty,$$

it follows from (8) that

$$(27) \quad g_n = \sum_{d|n} c_d \phi_{n/d}.$$

In fact, both of the limit processes, which lead to this evaluation of

$$\int_0^\pi g(t) \sin nt \, dt$$

or g_n , can be carried out "term-by-term," the legitimacy of the term-by-term integrations being assured by convergence in the mean (L^2) .

According to (26) and (27), the two functions, $f(t)$ and $g(t)$, have the same Fourier sine constants, f_n and g_n , on $(0, \pi)$. It follows therefore from the uniqueness theorem of Fourier sine series (L^2) that $f(t) = g(t)$ (almost everywhere). If this is compared with the one-to-one correspondence established by (23) and (24) between Hilbert spaces (22) and (3), it is seen that, in order to complete the proof of (III), only the assertion, (21), of (III'') remains to be verified.

The proof of (21) can be based on the fact used after (14), namely, on the circumstance that the Hermitian matrix (9) is the product $D^*(\Phi)D(\Phi)$. This implies that the matrix (9) is non-negative definite. Actually, since $D(\Phi)$ has a bounded inverse, $D^{-1}(\Phi)$, the bounded matrix (9) must be positive

definite, that is to say such that $\mu \geq 0$ can be refined to $\mu > 0$, if μ denotes the greatest lower bound attained by the Hermitian form belonging to the matrix (9), on the boundary of the unit sphere in Hilbert's space (Toeplitz). In particular,

$$\sum_{n=1}^{\infty} \sum_{m=1}^{\infty} c_n c_m^* \int_0^{\pi} \phi(nt) \phi^*(mt) dt = 0$$

holds only when $(c_1, c_2, \dots) = (0, 0, \dots)$. Since this double sum is the limit, as $k \rightarrow \infty$, of

$$\sum_{n=1}^k \sum_{m=1}^k c_n c_m^* \int_0^{\pi} \phi(nt) \phi^*(mt) dt = \int_0^{\pi} |0 - \sum_{n=1}^k c_n \phi(nt)|^2 dt,$$

the assertion, (21), of (III'') follows.

This completes the proof of (III). The proof of (III bis) proceeds in the same way, if the preceding deduction is first applied to $f(t) - c_0$, that is, to the function which results if the constant term of the expansion (19 bis) is subtracted from the function $f(t)$.

In order to illustrate the nature of the limitations imposed on $\phi(t)$, it is instructive to consider the case in which the last assumption, (7), of (III) is relaxed to

$$(28) \quad \Phi(s) \neq 0, \text{ where } \sigma > 0,$$

but all the other assumptions of (III) are retained. It is clear from (ii), and from the proof of (III), that the assertions of (III) cannot then remain true to their full extent. What is interesting is that not even (III'), a statement much weaker than (III), is such as to allow the reduction of (7) to (28).

In order to see this, it is sufficient to choose (8) as follows: $\phi(t) = \sin 2t$. Then every ϕ_n except $\phi_2 = 1$ is 0. Hence, (5) becomes $\Phi(s) = 2^{-s}$, and so (6) and the relaxed form, (28), of (7) are satisfied. But (7) is violated (in fact $|1/\Phi(s)| = 2^{\sigma} \rightarrow \infty$ as $\sigma \rightarrow \infty$), and this alone vitiates the assertion of (III'). For, since $\phi(t) = \sin 2t$, the sequence (20) now becomes

$$(28 \text{ bis}) \quad \sin 2t, \sin 4t, \sin 6t, \dots,$$

a sequence which cannot be (L^2) -complete on $(0, \pi)$. In fact, this sequence is a (proper) subsequence of the sequence of all odd harmonics. But no such subsequence can be (L^2) -complete on $(0, \pi)$, since the sequence of all odd harmonics is an orthogonal sequence on $(0, \pi)$.

REFERENCES

- [1] H. Burkhardt, "Entwicklungen nach oscillierenden Functionen und Integration der Differentialgleichungen der mathematischen Physik," *Jahresbericht der Deutschen Mathematiker-Vereinigung*, vol. 10, part 2 (1908), pp. 905-906.
- [2] A. Töpler, *Grazer Mitteilungen*, 1872, pp. 66-112 (the complete reference is not available here); cf. [1].
- [3] O. Toeplitz, "Zur Theorie der Dirichletschen Reihen," *American Journal of Mathematics*, vol. 60 (1938), pp. 880-888.
- [4] A. Wintner, "Ueber die Spektra der Toeplitzschen D -Formen," *Monatshefte für Mathematik und Physik*, vol. 48 (1939), pp. 147-152.
- [5] ———, "Diophantine approximations and Hilbert's space," *American Journal of Mathematics*, vol. 66 (1944), pp. 564-578.

ON RIEMANN'S REDUCTION OF DIRICHLET SERIES TO POWER SERIES.*

By AUREL WINTNER.

Subject to provisos of convergence, the reduction in question connects

$$f(s) = \sum_{n=1}^{\infty} a_n/n^s \text{ and } F(r) = \sum_{n=1}^{\infty} a_n r^n$$

by the formal identity

$$\Gamma(s)f(s) = \int_0^{\infty} x^{s-1} F(e^{-x}) dx$$

(cf. [8] and, for references concerning proofs of convergence in case of general Dirichlet series, [4], p. 11).

The Tauberian aspects of this connection seem to have been neglected in recent literature. An exception is one of the Hardy-Littlewood proofs of the prime number theorem ([3], pp. 128-133).

Today, it is possible to develop a Tauberian theory of the transformation $F \rightarrow f$ in a manner leading to more or less complete results. Curiously enough, even results containing $F(r)$ alone (cf. (III) below) can be obtained from the representation of f as the Mellin transform of F .

1. First, a few primitive facts will have to be collected.

(i) *If a Dirichlet series*

$$(1) \quad f(s) = \sum_{n=1}^{\infty} a_n/n^s$$

is convergent in the half-plane $\sigma > \lambda$, then the corresponding power series

$$(2) \quad F(r) = \sum_{n=1}^{\infty} a_n r^n$$

is convergent when $r < 1$ and satisfies

$$(3) \quad F(r) = O(1-r)^{-c} \text{ if } c > \max(0, \lambda),$$

as $r \rightarrow 1$.

* Received February 14, 1947.

In fact, if $\lambda \geq 0$, then the convergence of (1) for $\sigma > \lambda$ means that

$$(4) \quad \sum_{m=1}^n a_m = O(n^{\lambda+\epsilon})$$

holds for every $\epsilon > 0$. This implies that (2) is convergent when $r < 1$ and that, since

$$(5) \quad \sum_{n=1}^{\infty} a_n r^n = (1-r) \sum_{n=1}^{\infty} \sum_{m=1}^n a_m r^n,$$

the function $F(r)$ is

$$(1-r) \sum_{n=1}^{\infty} O(n^{\lambda+\epsilon}) r^n = (1-r) O(1-r)^{-\lambda-\epsilon-1}$$

as $r \rightarrow 1$. This proves (3) under the assumption $\lambda \geq 0$. But this assumption is made superfluous by the wording of (3).

(ii) *If the abscissa of convergence of (1) is non-negative and finite, then the integral*

$$(6) \quad \int_0^{\infty} x^{s-1} F(e^{-x}) dx, \quad (\text{cf. (2)}),$$

which is, when $\sigma < 1$, improper at $x = 0$, is absolutely convergent within the half-plane of (not necessarily absolute) convergence of (1).

The half-planes are meant to be open.

If λ denotes the abscissa of convergence of (1), the assumptions of (ii) are $0 \leq \lambda < \infty$. According to (i), these imply that $F(r) = O(1-r)^{-\lambda-\epsilon}$ as $r \rightarrow 1-0$, where $\epsilon > 0$ is arbitrary. Since this means $F(e^{-x}) = O(x^{-\lambda-\epsilon})$ as $x \rightarrow +0$, it follows that the contribution of the range $0 < x \leq 1$ to (6) is absolutely convergent if the real part, σ , of s satisfies the inequality

$$(\sigma - 1) - (\lambda + \epsilon) > -1,$$

that is, if $\sigma > \lambda$. On the other hand, the contribution of the range $1 \leq x < \infty$ to (6) is absolutely convergent for every s , since

$$(7) \quad F(e^{-x}) = O(e^{-x}) \text{ as } x \rightarrow \infty,$$

by (2) (where $F(0) = 0$, hence $F(r) = O(r)$ as $r \rightarrow 0$). This proves (ii).

(iii) *If (1) is convergent when $\sigma > \lambda$, and if $\lambda \geq 0$, then*

$$(8) \quad f(s) = \int_0^{\infty} x^{s-1} F(e^{-x}) dx / \Gamma(s)$$

when $\sigma > \lambda$.

This is that connection between (1) and (2) on which Riemann's proofs of the functional equation of $\zeta(s)$ depend. His starting point is that (8), for $\sigma > 0$, is obvious from

$$(9) \quad \Gamma(s) = \int_0^{\infty} x^{s-1} e^{-x} dx, \quad \sigma > 0,$$

if (1) and (2) reduce to monomials, $f(s) = 1/n^s$ and $F(r) = r^n$. Hence, (8) is true, for reasons of distributivity, in the half-plane $\sigma > 0$, if (1) is a finite sum. This reduces the problem to one concerning the legitimacy of a term-by-term integration in an integral (which is improper at $x = \infty$ and, if $\sigma < 1$, at $x = 0$ also). This term-by-term integration can be justified from (7) and (3), if $\sigma > \max(0, \lambda)$.

The last step in this proof of (iii) is just a routine matter. Nevertheless, this, the traditional, verification of (iii) is by no means as straightforward as it appears. In fact, it thinks of the problem as one concerning the *division* of (6) by (9), whereas the actual problem is that of the *multiplication* of (1) by (9). Correspondingly, a proof of (iii) which is less artificial than the usual one and which, without additional labor, leads much further than (iii), is contained in a general lemma concerning multiplication. In fact, (1) can be written as a *convergent*, and (9) as an *absolutely convergent*, Mellin integral, and so nothing more is involved than the integral form of the Mertens-Stieltjes multiplication theorem on series.

2. An immediate consequence of (iii) is the following fact, which claims that, if (1) is convergent for $\sigma > 1$ and if the coefficient sequence a_1, a_2, \dots , has an "*A*-mean," then the sequence has a "*D*-mean" as well (and the latter is equal to the former):

(iv) *If the Dirichlet series (1) is convergent for $\sigma > 1$ and if the corresponding power series (2) (which, by (i), must then converge for $r < 1$) is such that*

$$(10) \quad (1 - \bar{r})F(r) \rightarrow l \text{ as } r \rightarrow 1$$

holds for some $l = \text{const.}$, then

$$(11) \quad \epsilon f(1 + \epsilon) \rightarrow l \text{ as } \epsilon \rightarrow 0.$$

The implication claimed by (iv) is well-known. It can be completed by the following *necessary condition* for (10) (if (1) is convergent when $\sigma > 1$):

(v) *If l is fixed, then, under the assumptions of (iv),*

$$(12) \quad \epsilon f(1 + \epsilon + it) \rightarrow 0 \text{ as } \epsilon \rightarrow 0, \text{ if } 0 < \pm t < \infty$$

(but (12) must be replaced by (11) at $t = 0$).

From the proof of the necessity of $\zeta(1 + it) \neq 0$ for the truth of the prime number theorem, standard is that particular case of (v) in which the (A)-assumption, (10), is strengthened to the corresponding (C, 1)-assumption,

$$(13) \quad (a_1 + \cdots + a_n)/n \rightarrow l \quad (n \rightarrow \infty).$$

Actually, both (iv) and (v) can be concluded from (iii).

In fact, the assumptions of (iii) are now satisfied if $\sigma > 1$. Hence, from (8),

$$(14) \quad \Gamma(1 + \epsilon + it)f(1 + \epsilon + it) = \int_0^\infty x^{\epsilon+it} F(e^{-x}) dx,$$

if t is real and $\epsilon > 0$. But the assumption (10) means that

$$(15) \quad F(e^{-x}) = lx^{-1} + o(x^{-1}) \text{ as } x \rightarrow 0.$$

It follows therefore from (7) that, as $\epsilon \rightarrow 0$, the integral on the right of (14) is

$$\int_0^1 x^{\epsilon+it} lx^{-1} dx + O(1) = l \int_0^1 x^{-1+\epsilon+it} dx + O(1) = l/(\epsilon + it) + O(1).$$

Consequently, if (14) is multiplied by ϵ ,

$$\epsilon \Gamma(1 + \epsilon + it)f(1 + \epsilon + it) = l\epsilon/(\epsilon + it) + O(\epsilon).$$

This relation proves the assertions, (11) and (12), of (iv) and (v), since $\Gamma(1 + \epsilon) \rightarrow 1$ if $t = 0$ and $\Gamma(1 + \epsilon + it) \rightarrow \Gamma(1 + it) \neq 0$ if $t \neq 0$.

3. According to an "elementary" theorem of Hardy and Littlewood, proved very simply by Karamata [6], the Tauberian restriction

$$(16) \quad a_n \geq 0$$

is sufficient in order that the converse of Frobenius' unconditional implication, (13) \rightarrow (10), be true. In other words,

$$(17) \quad (10) \text{ and } (16) \text{ imply } (13).$$

On the other hand,

(11) and (16) do not imply (13)

or, what (in view of (17)) is the same thing,

(18) (11) and (16) do not imply (10).

In fact, (18) is clear from Dedekind's example,

$$(19) \quad a_n = n \text{ if } n = 2^k \text{ and } a_n = 0 \text{ if } n \neq 2^k, \quad (k = 0, 1, \dots),$$

(cf. [1]). What is responsible for this situation is precisely (v). For, in order to prove (18), it is sufficient to ascertain that the convergence of (1) for $\sigma > 1$ and the assumptions (11), (16) together do not imply the assertion, (12), of (v). But (16) is satisfied in the case (19), and the series (1) becomes

$$f(s) = \sum_{k=0}^{\infty} 2^k / 2^{ks} = \sum_{k=1}^{\infty} (2^{-s+1})^k,$$

a geometric progression which converges when $\sigma > 1$. However, since $f(s)$ has the period $2\pi i / \log 2$ and possesses a simple pole at $s = 1$, the necessary condition expressed by (12) is violated at $t = \pm 2\pi i n / \log 2$, even though (11) is satisfied.

In order to satisfy (12), Ikehara's theorem, which retains (16) and the convergence of (1) for $\sigma > 1$, assumes both (11) and (12), and somewhat more; substantially, the existence of a continuous boundary function for

$$f(s) - l/(s-1)$$

on the line $\sigma = 1$ (cf., e. g., [7]).

Since the local behavior of $l/(s-1)$ on $\sigma = 1$ is the same as that of $l\zeta(s)$, Ikehara's theorem can be formulated as follows:

If the Dirichlet series

$$(20) \quad f^*(s) = f(s) - l\zeta(s) = \sum_{n=1}^{\infty} (a_n - l)/n^s$$

converges in the half-plane $\sigma > 1$ to a function which attains continuous boundary values on the line $\sigma = 1$, then (16) is sufficient for (13).

4. This theorem will now be refined so as to replace Ikehara's assertion, (13), by (10) and Ikehara's Tauberian restriction, (16), by a more inclusive assumption, to be placed on the *function* (2) (which represents *averages* of a_1, a_2, \dots), rather than on the *individual coefficients* a_n , as follows:

(vi) *Let the Dirichlet series (1) be convergent in the half-plane $\sigma > 1$.*

Then the Abelian relation (10) holds for the power series (2) whenever the coefficients a_n satisfy the following pair of conditions:

(α) $F(r) = \sum_{n=1}^{\infty} a_n r^n$ is a monotone function of r as $r \rightarrow 1$ (i. e., ultimately), and

(β) the function (20), where $\sigma > 1$, goes over into a continuous boundary function on the line $\sigma = 1$; a requirement which can be relaxed to the assumption that

(β bis) for every positive $T < \infty$, the functions

$$(21) \quad f_{\sigma}^*(t) = f^*(\sigma + it), \quad -\infty < t < \infty,$$

defined by (20) for $\sigma > 1$, converge, as $\sigma \rightarrow 1$, in the mean of the function space $(L) = (L^1)$ on $(-T, T)$ (that is,

$$(22) \quad \int_{-T}^T |f^*(1 + \delta + it) - f^*(1 + \eta + it)| dt \rightarrow 0 \text{ as } \delta^2 + \eta^2 \rightarrow 0,$$

where $\delta > 0$ is independent of $\eta > 0$).

Remark. In order that (22) be satisfied, it is sufficient, but not necessary, that the functions (21) of t be majorized by an L -integrable function of t on $(-T, T)$, as $\sigma \rightarrow 1$.

Since (16) is sufficient for (α) in (vi), it follows from (17) that Ikehara's theorem is contained in (vi). The converse inference is not possible in any sense. In fact, a domain which is reached by (vi) but cannot be reached by Ikehara's theorem is exemplified by so primitive a case as

$$(23) \quad a_n = (-1)^{n+1} n.$$

For, since $a_n \neq o(n)$ in this case, the assertion, (13), of Ikehara's theorem is surely false. But the assertion, (10), of (vi) is true; and the point is that this can be concluded from (vi) itself. In fact, (1) converges, in the case (23), in the half-plane $\sigma > 1$ to a function, $f(s)$, which is an entire function, hence such as to satisfy (β) (with $l=0$), and (α) is fulfilled, since (2) becomes

$$F(r) = \sum_{n=1}^{\infty} (-1)^{n+1} n r^n = r/(1+r)^2 \rightarrow \frac{1}{4} \quad (1 > r \rightarrow 1).$$

5. In order to prove (vi), suppose first only that (1) is convergent

when $\sigma > 1$. This means that the assumption of (iii) is satisfied by $\lambda = 1$. Hence, (8) is applicable if the real part of $s = \sigma + it$ exceeds 1. Accordingly,

$$(25) \quad f(s)\Gamma(s) = \int_{-\infty}^{\infty} e^{-sy} F(\exp - e^{-y}) dy,$$

if x is replaced by e^{-y} in (8).

In view of (i), the integral (25) is absolutely convergent. Hence, if $\phi(t)$ is any L -integrable function on a finite interval $(-T, T)$, then

$$(26) \quad \int_{-T}^T \int_{-\infty}^{\infty} |\phi(t) e^{-sy} F(\exp - e^{-y})| dy dt < \infty.$$

Consequently, from (25), and by Fubini's theorem,

$$(27) \quad \int_{-2T}^{2T} \phi(t) f(s) \Gamma(s) dt = \int_{-\infty}^{\infty} F(\exp - e^{-y}) e^{-\sigma y} \int_{-2T}^{2T} e^{-ity} \phi(t) dt dy,$$

if T is replaced by $2T$.

As in the standard proof of Ikehara's theorem (cf., e. g., [7]), choose

$$\phi(t) = \frac{1}{2} (1 - \frac{1}{2} |t|/T) e^{ixt},$$

where x is a real number. Then the interior integral on the right of (27) becomes

$$\frac{1}{2} \int_{-2T}^{2T} e^{i(x-y)t} (1 - \frac{1}{2} |t|/T) dt = T \Phi(Tx - Ty),$$

if Φ is an abbreviation for the function

$$(28) \quad \Phi(u) = (\sin u)^2 / u^2.$$

Hence, (27) reduces to

$$(29) \quad \begin{aligned} & \frac{1}{2} \int_{-2T}^{2T} (1 - \frac{1}{2} |t|/T) e^{ixt} f(\sigma + it) \Gamma(\sigma + it) dt \\ &= T \int_{-\infty}^{\infty} F(\exp - e^{-y}) e^{-\sigma y} \Phi(Tx - Ty) dy, \end{aligned}$$

where $\sigma > 1$.

Clearly, (29) remains true if f and F are replaced by f^* and F^* respectively, where, corresponding to (1), (2) and (20),

$$(30) \quad F^*(r) = \sum_{n=1}^{\infty} (a_n - l)r^n = F(r) - lr/(1-r).$$

Let (31) denote the identity which results from (29) upon the replacement $(f, F) \rightarrow (f^*, F^*)$.

In view of the completeness of the L -space, the last assumption, (β bis), of (vi) implies that, as $\sigma \rightarrow 1$, the functions (21) tend in the mean (L) to an L -integrable function, say $f_0(t)$, on every finite t -interval. Furthermore, this limit process can be carried out beneath the integral sign of the integral on the left of (31) (the convergence being "strong" with reference to the L -space). Finally, since (31) is true when $\sigma > 1$, the expression on the right of (31) must tend, as $\sigma \rightarrow 1$, to a finite limit, the latter being the limit of the expression on the left. Accordingly, if

$$(32) \quad g_T(t) = \frac{1}{2}(1 - \frac{1}{2}|t|/T)f_0(t)\Gamma(1+it)$$

and

$$(33) \quad \phi_T(x) = T \lim_{\sigma \rightarrow 1} \int_{-\infty}^{\infty} F^*(\exp - e^{-y}) e^{-\sigma y} \Phi(Tx - Ty) dy,$$

the result of the limit process in (31) is

$$\int_{-2T}^{2T} e^{ixt} g_T(t) dt = \phi_T(x),$$

along with the existence of (33) (as a finite limit) for every real x and for every $T > 0$. Finally, since $f_0(t)$, and therefore the function (32) of t , is L -integrable on every finite t -interval, the last formula line shows that, in view of the Riemann-Lebesgue lemma,

$$(34) \quad \phi_T(x) \rightarrow 0 \text{ as } x \rightarrow \pm \infty.$$

6. Condition (α) of (vi) has not been used thus far. Hence (if f^* , F^* are replaced by f , F), the result of the preceding deduction can be formulated as follows:

(vii) *If the Dirichlet series (1) is convergent when $\sigma > 1$ and satisfies, for a fixed $T > 0$, the condition*

$$(35) \quad \int_{-2T}^{2T} |f(1 + \delta + it) - f(1 + \eta + it)| dt \rightarrow 0 \text{ as } (\delta, \eta) \rightarrow (0, 0),$$

then the Abelian value,

$$(36) \quad \lim_{\epsilon \rightarrow 0} \int_{-\infty}^{\infty} e^{-\epsilon y} F(\exp - e^{-y}) e^{-y} \Phi(Tx - Ty) dy,$$

of the improper integral

$$(37) \quad \int_{-\infty}^{\infty} F(\exp - e^{-y}) e^{-y} \Phi(Tx - Ty) dy,$$

where F, Φ are defined by (2), (28), exists, as a finite limit, for every real x , and

$$(38) \quad (36) \text{ tends to } 0 \text{ as } x \rightarrow \pm \infty.$$

The convergence of (37) is not claimed. The transition from (36) to (37) would correspond to one from the (A)-summability of a series $\sum c_n$ to its convergence, that is, from

$$(36') \quad \sum_{n=1}^{\infty} c_n e^{-\epsilon n} \rightarrow C \text{ as } \epsilon \rightarrow 0$$

to

$$(37') \quad \sum_{n=1}^N c_n \rightarrow C \text{ as } N \rightarrow \infty,$$

which, according to Tauber's theorem, is justified if (and only if)

$$(39') \quad \sum_{n=1}^N n c_n = o(N).$$

Correspondingly, by the integral analogue of Tauber's theorem,

(vii bis) (36) in (vii) can be replaced by (37) if (and only if)

$$(39) \quad \int_0^R y F(\exp - e^{-y}) e^{-y} \Phi(Tx - Ty) dy = o(R) \text{ as } R \rightarrow \infty,$$

where x is fixed (as is T).

Strictly speaking, (39) belongs to those modifications of (36), (37) in which the lower limit of integration, instead of being $y = -\infty$, is $y = 0$, as in (39). Actually, the contribution of the half-line $y < 0$ to (36), (37)

is immaterial, since, if $r = \exp - e^{-y}$, this half-line represents an r -interval ending *before* the crucial point, $r = 1$, in (2).

A condition more superficial than (39) is

$$(40) \quad F(r) \geq 0$$

(as $r \rightarrow 1$). In fact, by (28),

$$(41) \quad \Phi \geq 0,$$

and so it is clear that (36) can be replaced by (37) if (but not only if) (40) is satisfied.

7. After this interruption of the proof of (vi), consider again (20), (30), instead of (1), (2), hence (33), instead of (36).

If the analogue of Tauber's condition, (39'), can be assumed, then (33) simplifies to

$$(42) \quad \phi_T(x) = T \int_{-\infty}^{\infty} F^*(\exp - e^{-y}) e^{-y} \Phi(Tx - Ty) dy,$$

and this would admit a direct application of (α) along the lines of the "Tauberian" argument of Karamata-Wiener. Fortunately, the circumstance that the integral (42) must be confined to its Abelian form, (33), will not lead to difficulties.

First, the assumption (α) is that the function (2) is monotone when r is close enough to its limiting value, 1. After a suitable alteration of the first coefficient, a_1 , of (2), an alteration which clearly has no influence on the assertion and the remaining assumptions of (vi), it can be assumed that $F(r)$ is monotone from the beginning, that is, on the whole interval $0 < r < 1$. Hence, $F(r)$ will be strictly monotone, and will not therefore become 0 more than once, when r varies from $r = 0$ to $r = 1$. After an additional alteration of a_1 , it can be assumed that $F(r)$ does not become 0 at all (except at $r = 0$), and so, since $F(r)$ can be replaced by $-F(r)$, that

$$(43) \quad F(r) > 0 \text{ if } 0 < r < 1.$$

Two cases are possible, according as the monotone function $F(r)$ is decreasing or increasing. In the first case, (43) implies that $F(r)$ is bounded as $r \rightarrow 1$ and must therefore tend to a finite limit, $F(1 - 0)$. Since this implies what is claimed in (vi) (and much more), it is sufficient to consider the second case, where

$$(44) \quad F(r') < F(r'') \text{ if } r' < r''.$$

It should be noted that, in the above normalization, the formal "residue" occurring in (20), (30) cannot be negative, since (2) and (43) imply that

$$(45) \quad l \geq 0.$$

Incidentally, the proof of (vi) admits of a slight simplification if the sign of equality holds in (45), since (30) then reduces to (2). But the assertion of (vi) is not more evident in this limiting case than it is when $l \neq 0$. Actually, the contribution of the trivial second term in (30) or, if

$$(46) \quad F_0(r) = \sum_{n=1}^{\infty} r^n = r/(1-r),$$

in

$$(47) \quad F^*(r) = F(r) - lF_0(r),$$

can be isolated in any case.

8. According to (47), the contribution of (46) to the integral following the lim-sign in (33) is

$$\int_{-\infty}^{\infty} E_{\sigma}(y) \Phi(Tx - Ty) dy,$$

where

$$E_{\sigma}(y) = \exp(-e^{-y} - \sigma y) / (1 - \exp - e^{-y})$$

and $\sigma > 1$. If $\sigma \rightarrow 1$, this quotient tends to

$$(48) \quad E(y) = \exp(-e^{-y} - y) / (1 - \exp - e^{-y}),$$

and therefore the last integral to

$$(49) \quad \int_{-\infty}^{\infty} E(y) \Phi(Tx - Ty) dy.$$

In fact, it is clear from (48) that

$$E(y) > 0 \text{ for } -\infty < y < \infty$$

and, as $y \rightarrow \infty$,

$$E(y) \sim \exp(-e^{-y} - y) / e^{-y} = \exp(-e^{-y}) \rightarrow 1, \quad (y \rightarrow \infty),$$

finally, as $y \rightarrow -\infty$,

$$E(y) = o(1) / (1 - o(1)) = o(1), \quad (y \rightarrow -\infty).$$

On the other hand, from (28),

$$\Phi(Tx - Ty) = O(y^{-2}) \text{ as } y \rightarrow \pm \infty,$$

if x and T are fixed. But the four last formula lines imply the convergence of the integral (49), and the legitimacy of the limit process ($\sigma \rightarrow 1$) which led to (49) is just as obvious.

Accordingly, the contribution of the second term on the right of (47) to the limit on the right of (33) is $-l$ times the value of the (convergent) integral (49). Hence, (33) can be decomposed into

$$(50) \quad \phi_T(x) = T \lim_{\sigma \rightarrow 1} \int_{-\infty}^{\infty} e^{-\sigma y} F(\exp - e^{-y}) \Phi(Tx - Ty) dy - lT\psi_T(x),$$

where, according to (49) and (48),

$$(51) \quad \psi_T(x) = \int_{-\infty}^{\infty} e^{-Y-y} (1 - e^{-Y})^{-1} \Phi(Tx - Ty) dy, \quad Y = e^{-y}.$$

Since the limit occurring on the right of (50) exists, it follows from (43) and (41) that (50) can be replaced by

$$(52) \quad \phi_T(x) = T \int_{-\infty}^{\infty} e^{-y} F(\exp - e^{-y}) \Phi(Tx - Ty) dy - lT\psi_T(x).$$

In fact, this reduction of (50) depends only on the trivial criterion, (40), mentioned after the necessary and sufficient condition, (vii bis).

Needless to say, what this reduction actually accomplishes is that the representation (33) can be improved to (42).

9. It is seen from (51), (28), and from the four formula lines preceding (50), that

$$\lim_{x \rightarrow \infty} \psi_T(x) = \int_{-\infty}^{\infty} \Phi(Ty) dy.$$

The last integral is

$$\int_{-\infty}^{\infty} \Phi(Ty) dy = \int_{-\infty}^{\infty} \Phi(y) dy / T = \pi / T,$$

by (28). According to (34),

$$\lim_{x \rightarrow \infty} \phi_T(x) = 0.$$

Finally, (52) can be written in the form

$$\phi_T(x) + lT\psi_T(x) = \mu_T(x),$$

where

$$(53) \quad \mu_T(x) = \int_{-\infty}^{\infty} e^{-y/T} F(\exp - e^{-y/T}) \Phi(Tx - y) dy.$$

The four formula lines which precede (53) imply that

$$(54) \quad \lim_{x \rightarrow \infty} \mu_T(x) = \pi l, \text{ where } \pi = \int_{-\infty}^{\infty} \Phi(y) dy,$$

is an identity in T . On the other hand, the assertion of (vi) is (10) or, what is the same thing,

$$(55) \quad \lim_{u \rightarrow \infty} e^{-u} F(\exp - e^{-u}) = l.$$

Hence, the proof of (vi) is complete if it is verified that (53), (54) and (41) imply (55) by virtue of *both* assumptions (43), (44), the *second* of which has not been used thus far. But the proof of this implication is exactly the same as in the corresponding final step in the proof of Ikehara's theorem (namely, steps 1)-2) in [7], pp. 526-527) and will therefore be omitted.

10. The theorem just proved, (vi), implies the following:

(viii) *If a Dirichlet series (1) is convergent for $\sigma > 1$ and, for some constant l , the corresponding function (20) satisfies the boundary condition (22) for every positive $T < \infty$, then*

$$(56) \quad (1-r) \sum_{n=1}^{\infty} a_n r^n \rightarrow l \text{ as } r \rightarrow 1$$

holds whenever the coefficients of (1) satisfy the estimate

$$(57) \quad \sum_{n=1}^x a_n = O(x).$$

Clearly, the restriction

$$(58) \quad a_n = O(1)$$

is sufficient for (57). Under this restriction, Ikehara's theorem supplies (13), instead of just (56). Nevertheless, it is seen from (17) that (13) can be concluded from (viii) also, if the Tauberian conditions of (viii) are strengthened to those of Ikehara's theorem. Conversely, if the latter restrictions are replaced by those of (viii), then (56) can hold when (13) is false. This is shown by the example (23), which satisfies (57) and (22) (with $l = 0$ in (20)).

The true theorem is, however, one which, instead of the partial sums

$$(59) \quad s_n = \sum_{m=1}^n a_m$$

occurring in (viii), involves only the sums

$$(60) \quad t_n = \sum_{m=1}^n m a_m,$$

which occur in Tauber's own theorem. In fact, from (59) and (60),

$$t_n = n s_n - \sum_{m=1}^{n-1} s_m.$$

Hence,

$$\text{if } s_n = O(n), \text{ then } t_n = O(n^2).$$

Consequently, (viii) is contained in the following theorem:

(ix) *In (viii), the coefficient restrictions can be replaced by*

$$(61) \quad \sum_{n=1}^{\infty} n a_n = O_L(x^2)$$

(in particular, (16) can be relaxed to its averaged form,

$$(62) \quad \sum_{n=1}^{\infty} n a_n \geq 0,$$

and (57) to

$$(63) \quad \sum_{n=1}^{\infty} n a_n = O(x^2),$$

the corresponding averaged condition).

It is readily verified that, by adding a constant to every a_n , one can reduce the sufficiency of (61) to that of (62). Hence, in order to prove both (viii) and (ix), it is sufficient to show that condition (α) of (vi) is satisfied if (62) is assumed. In other words, it is sufficient to ascertain that the derivative of (2) is non-negative if the sums t_n , defined by (60), are non-negative. But this is obvious, since the derivative of (2) is

$$\sum_{n=1}^{\infty} n a_n r^{n-1} = \sum_{n=1}^{\infty} (t_n - t_{n-1}) r^{n-1} = (1-r) \sum_{n=1}^{\infty} t_n r^{n-1}.$$

11. The rôle of (61) in (ix) is that of supplying a sufficient condition for the first assumption, (α) , of (vi). It will now be shown that a sufficient condition for the remaining assumption, $(\beta \text{ bis})$, of (vi) (that is, for (22), where $0 < T < \infty$), is supplied by

$$(64) \quad \int_0^{1-0} (1-r) |F^*(r)|^2 dr < \infty.$$

(x) Let

$$(65) \quad f(s) = \sum_{n=1}^{\infty} a_n/n^s, \text{ hence } f^*(s) = \sum_{n=1}^{\infty} (a_n - l)n^s,$$

be convergent for $\sigma > 1$ (and, therefore,

$$(66) \quad F(r) = \sum_{n=1}^{\infty} a_n r^n \text{ and } F^*(r) = \sum_{n=1}^{\infty} (a_n - l)r^n$$

for $r < 1$). Then (61) and (64), respectively, are sufficient for the assumptions (α) and (β bis) of (vi) and imply, therefore, that

$$(67) \quad (1-r)F(r) \rightarrow l, \text{ i. e., } F^*(r) = o(1-r)^{-1}, \text{ as } r \rightarrow 1.$$

There is something unsatisfactory in the structure of the last conclusion, since both the assumptions, (61) and (64), and the assertion, (67), involve only the power series, (66). Thus the sole part played by the Dirichlet series, (65), consists in their convergence in the half-plane $\sigma > 1$. This, an assumption *not* implied by the convergence of the power series for $r < 1$, enters, of course, via (vi).

This anomaly can be disposed of, since (vi) can be generalized as follows:

$$(I) \quad \text{Let } F(r) = \sum_{n=1}^{\infty} a_n r^n \text{ and}$$

$$(68) \quad \phi(s) = \int_0^{\infty} x^{s-1} F(e^{-x}) dx$$

be convergent for $r < 1$ and $\sigma > 1$ respectively. Suppose that

$$(69) \quad F(r) \text{ is monotone as } r \rightarrow 1-0$$

(for instance, that (58) is satisfied) and that the function

$$(70) \quad \phi^*(s) = \phi(s) - l\zeta(s)$$

or, equivalently, the function

$$(71) \quad \phi^*(s) = \phi(s) - l/(s-1),$$

where l is a constant, satisfies the boundary condition

$$(72) \quad \int_{-T}^T |\phi^*(1 + \delta + it) - \phi^*(1 + \eta + it)| dt \rightarrow 0 \text{ as } (\delta, \eta) \rightarrow (0, 0),$$

for every fixed $T < \infty$. Then, as $r \rightarrow 1$,

$$(73) \quad (1 - r)F(r) \rightarrow l, \text{ i. e., } \sum_{n=1}^{\infty} (a_n - l)r^n = o(1 - r)^{-1}.$$

The point is that this theorem, (I), includes Dirichlet series (65) which have no half-planes of convergence. Nevertheless, the proof of (I) is the same as that of (vi). In fact, if $f^*(s)$, instead of being defined by (20), is defined by

$$(74) \quad f^*(s) = \phi^*(s)/\Gamma(s), \quad (\sigma > 1),$$

where $\phi^*(s)$ is the function (70), then, since $\Gamma(s)$ is continuous and distinct from 0 on the closed half-plane $\sigma \geq 1$, the assumption (72) of (I) becomes equivalent to the assumption (β bis) of (vi), whereas (69) represents (α) in (vi). But this suffices for the proof of (vi). For, on the other hand, (iii) implies that, if the Dirichlet series (1) or (20) is convergent for $\sigma > 1$, then (74) holds by virtue of (68) and (70), and, on the other hand, only the functions (68), (70), (74), rather than the two series (65), have been used in the proof of (vi).

12. This reduces the problem of (x) to one involving the power series (66) only. In fact, (x) can be reduced to (I) and to the following criterion:

(II) If a power series

$$(75) \quad G(r) = \sum_{n=1}^{\infty} b_n r^n$$

and the corresponding Mellin transform

$$(76) \quad \psi(s) = \int_0^{\infty} x^{s-1} G(e^{-x}) dx$$

are convergent for $r < 1$ and $\sigma > 1$ respectively, and if

$$(77) \quad \int_0^{1-0} (1-r) |G(r)|^2 dr < \infty,$$

then the function $\psi(s) = \psi(\sigma + it)$, where $\sigma > 1$, goes over into a locally

L-integrable boundary function on the line $\sigma=1$ (that is, there exists a measurable function $\psi_0(t)$, $-\infty < t < \infty$, satisfying

$$(78) \quad \int_{-T}^T |\psi(1+\epsilon+it) - \psi_0(t)| dt \rightarrow 0 \text{ as } \epsilon \rightarrow 0,$$

where $T < \infty$ is arbitrary).

Actually, the convergence of (76) in the half-plane $\sigma > 1$ need not be required, since even the *absolute* convergence of (76), for $\sigma > 1$, is implied by (77). In fact, (76) is convergent in the half-plane $\sigma > 1$ if and only if the improper integral

$$\int_0^{1-\epsilon} (\log r)^\epsilon G(r) dr/r$$

is convergent for every $\epsilon > 0$. Hence, (76) is absolutely convergent in the half-plane $\sigma > 1$ if and only if

$$\int_0^1 |(1-r)^\epsilon G(r)| dr < \infty \text{ when } \epsilon > 0.$$

But Schwarz's inequality shows that the last integral is majorized by the square root of JG_ϵ , where J denotes the integral (77) and

$$G_\epsilon = \int_0^1 |(1-r)^{-\frac{1}{2}+\epsilon}|^2 dr = \int_0^1 (1-r)^{-1+2\epsilon} dr < \infty.$$

Clearly, (78) implies that

$$(79) \quad \int_{-T}^T |\psi(1+\delta+it) - \psi(1+\eta+it)| dt \rightarrow 0 \text{ as } (\delta, \eta) \rightarrow (0, 0),$$

i. e., that (72) is satisfied for $\phi^* = \psi$. Hence, if (II) is granted, (I) supplies the following Tauberian theorem, which involves nothing but a power series:

(III) If $r < 1$, let

$$(80) \quad P(r) = \sum_{n=1}^{\infty} c_n r^n$$

be a convergent power series satisfying the integral condition

$$(81) \quad \int_0^{1-0} (1-r) |P(r)|^2 dr < \infty,$$

and suppose that, as $r \rightarrow 1$,

$$(82) \quad P(r) \text{ is a monotone function}$$

(which is sure to be the case if

$$(82^*) \quad \sum_{n=1}^x nc_n = O_L(x^2) \text{ as } x \rightarrow \infty;$$

for instance, if

$$(82') \quad \sum_{n=1}^x nc_n \geq 0$$

holds from a certain x onward). Then

$$(83) \quad P(r) = o(1-r)^{-1} \text{ as } r \rightarrow 1.$$

Clearly, (x) is a corollary of (III). Since (I) and (II) imply (III), and since (I) has already been verified, it follows that only (II) remains to be proved.

13. The assertion of (II) is the existence of a measurable function $\psi_0(t)$, $-\infty < t < \infty$, satisfying (78), where T is arbitrary ($< \infty$). In view of the completeness of the L -space on a t -interval, the existence of such a $\psi_0(t)$ is equivalent to (79). On the other hand, since $T < \infty$, it follows from Schwarz's inequality that (79) is sure to be true if $(79^2)_T$ is true, where $(79^2)_T$ denotes the relation which results if the function integrated in (79) is replaced by its square. But $(79^2)_T$, where $T < \infty$, is certainly true if $(79^2)_\infty$ is true (with the understanding that $\psi(\sigma + it)$ should be of class (L^2) on the line $-\infty < t < \infty$, if $\sigma > 1$ is fixed). Finally, Plancherel's theorem states that $(79^2)_\infty$ is true if and only if

$$(84) \quad \int_{-\infty}^{\infty} |\Psi_\delta(u) - \Psi_\eta(u)|^2 du \rightarrow 0 \text{ as } (\delta, \eta) \rightarrow (0, 0),$$

where (in the (L^2) -sense)

$$(85) \quad 2\pi\Psi_\epsilon(u) \sim \int_{-\infty}^{\infty} \psi(1 + \epsilon + it)e^{itu}dt$$

and, for every $\epsilon > 0$,

$$(86) \quad \int_{-\infty}^{\infty} |\psi(1 + \epsilon + it)|^2 dt < \infty, \text{ i. e., } \int_{-\infty}^{\infty} |\Psi_{\epsilon}(u)|^2 du < \infty.$$

Consequently, more than (II) will be proved if it is verified that the assumptions of (II) imply (86) and (84), where Ψ_{ϵ} is defined by (85) in virtue of (86).

To this end, put $s = 1 + \epsilon + it$ and $x = e^{-u}$ in (76). Then (76) appears in the form

$$\psi(1 + \epsilon + it) = \int_{-\infty}^{\infty} e^{-(1+\epsilon)u} G(\exp - e^{-u}) e^{-itu} du.$$

As verified after (II), this integral is absolutely convergent by virtue of the assumption (77) of (II). Since (77) implies that the function

$$e^{-(1+\epsilon)u} G(\exp - e^{-u})$$

is of class (L^2) (cf. below) and since, being analytic, it is of bounded variation locally, Fourier's inversion is legitimate. This means that the integral on the right of (85) is convergent (beforehand, as a "principal" integral in Cauchy's sense), and is equal to

$$2\pi e^{-(1+\epsilon)u} G(\exp - e^{-u}).$$

But now it is clear that (in view of the absolute convergence of (75) and (76) for $r < 1$ and $\sigma > 1$, respectively) the condition (86) is satisfied, and that the function on the left of (85) is precisely the function in the last formula line; in fact, the Fourier transform (L^2) of a function of class (L^2) is (L^2) -unique.

It also follows that (85), the only relation which remains to be verified, is equivalent to

$$\int_{-\infty}^{\infty} |(e^{-\delta u} - e^{-\eta u}) e^{-u} G(\exp - e^{-u})|^2 du \rightarrow 0 \text{ as } (\delta, \eta) \rightarrow (0, 0).$$

Since this condition is identical with

$$\int_0^{\infty} x |(x^{\delta} - x^{\eta}) G(e^{-x})|^2 dx \rightarrow 0 \text{ as } (\delta, \eta) \rightarrow (0, 0),$$

it is satisfied if

$$\int_0^{\infty} x |G(e^{-x})|^2 dx < \infty, \text{ i. e., } \int_0^1 (-\log r) |G(r)|^2 dr/r < \infty$$

($r = e^{-x}$). But, since $-\log r \sim 1 - r$ as $r \rightarrow 1$, the convergence of the last integral is equivalent to (76) (there is no trouble at $r = 0$, since, according to (75),

$$(-\log r) |G(r)|^2/r = (-\log r) O(r^2)/r \text{ as } r \rightarrow 0,$$

and this is $O(-r \log r) \rightarrow 0$).

14. This proves (II), and therefore all theorems formulated above. Actually, the proof is such as to lead, without additional effort, to certain variants of these theorems. This is illustrated by the following extension of (III):

(IV) *If $\lambda > 1$ and $\mu = \lambda/(\lambda - 1)$, then (III) remains true when its assumption (81) is generalized to*

$$(88) \quad \int_0^{1-0} (1-r)^{\mu/\lambda} |P(r)|^\mu dr < \infty$$

(which becomes (81) when $\lambda = 2 = \mu$).

In order to see this, it is sufficient to observe that the preceding application of Plancherel's theorem can be replaced by that of its Hölder form (Young, Hausdorff, Titchmarsh). Then all that remains to be ascertained is that, just as in the particular case (81), the (absolute) convergence of (76) for $\sigma > 1$ is always implied by (88), where $P = G$. Thus it is seen from the formula line following (77), where $\sigma = 1 + \epsilon$ (and $G = P$), that it is sufficient to prove

$$\int_0^{1-0} |\log r|^\epsilon |P(r)| dr/r < \infty, \text{ i. e., } \int_0^{1-0} (1-r)^\epsilon |P(r)| dr < \infty,$$

for every $\epsilon > 0$. But this follows from (88) in the same way as in the above case, $\lambda = 2 = \mu$, of Schwarz's inequality.

In view of the classical case, (β), of (β bis) in (vi), the following limiting form of (IV) is of particular interest:

(V) *Theorem (III) remains true if its assumption (81) is replaced by*

$$(89) \quad \int_0^{1-0} |P(r)| dr < \infty$$

(which is (88) with $\mu = 1 + 0$, $\lambda = \mu/(\mu - 1) = \infty$); in fact, (89) implies that

$$(90) \quad \psi(s) = \int_0^{\infty} x^{s-1} P(e^{-x}) dx$$

is absolutely convergent at $s = 1$, and so the analytic function defined by (90) in the half-plane $\sigma > 1$ goes over into a continuous boundary function on the line $\sigma = 1$.

It is true that, in the limiting case (89) of (88), the (L^p) -theory of Fourier transforms does not apply. But such a theory is not needed now, since what is required by (β bis) in (vi) is trivial from the continuity of (90) on the closed half-plane $\sigma \geq 1$. Hence, (V) follows from the proof of (vi) for the same reason as (x) did.

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES

-
- [1] R. Dedekind, Supplement IX, pp. 385-386 to the 3rd ed. (1879) of Dirichlet's *Vorlesungen über Zahlentheorie*.
 - [2] Hardy, G. H., *The Mathematical Work of Ramanujan*, Princeton, 1936.
 - [3] ——— and J. E. Littlewood, "Contributions to the theory of the Riemann zeta function and the theory of the distribution of primes, *Acta Mathematica*, vol. 41 (1918), pp. 119-196.
 - [4] ——— and M. Riesz, *The General Theory of Dirichlet's Series*, Cambridge Tracts, no. 18, 1915.
 - [5] A. E. Ingham, *The Distribution of Prime Numbers*, Cambridge Tracts, no. 30, 1932.
 - [6] J. Karamata, "Ueber die Hardy-Littlewoodschen Umkehrungen des Abelschen Stetigkeitssatzes," *Mathematische Zeitschrift*, vol. 32 (1930), pp. 319-320.
 - [7] E. Landau, "Ueber Dirichletsche Reihen," *Göttinger Nachrichten*, 1932, pp. 525-527.
 - [8] B. Riemann, *Gesammelte mathematische Werke*, 1876, pp. 136-138.

ON SKEW-METRIC SPACES AND FUNCTION GROUPS.*

By HWA-CHUNG LEE.

Introduction. In two previous papers¹ we have studied the even-dimensional spaces equipped with a nonsingular skewsymmetric fundamental tensor. We shall extend the consideration of such spaces to the cases where the fundamental tensor is not necessarily nonsingular, and the dimensionality may also be odd. We shall generalise the study of function groups to the *flat* spaces of this nature, and discover that the "group spaces" of function groups are natural examples of such flat spaces.

Let L_m be an m -dimensional space endowed with a skewsymmetric *covariant* fundamental tensor $a_{\alpha\beta}(x)$,² not necessarily nonsingular, whose components in each coordinate system are (differentiable) functions of the coordinates x^1, \dots, x^m . The rank of the matrix $a_{\alpha\beta}$ being necessarily even, let it be $2n$, which we also call the *rank* of L_m .

The skewsymmetric quantity

$$(1) \quad K_{\alpha\beta\gamma} = \frac{\partial a_{\beta\gamma}}{\partial x^\alpha} + \frac{\partial a_{\gamma\alpha}}{\partial x^\beta} + \frac{\partial a_{\alpha\beta}}{\partial x^\gamma}$$

is a covariant tensor, which we call the *curvature tensor* of L_m . If this curvature tensor vanishes, L_m is said to be *flat*. For a flat L_m , we may use results in the Pfaff Problem to prove the existence of a special coordinate system in which the exterior form $a_{\alpha\beta}[dx^\alpha dx^\beta]$ reduces to³

$$2[dx^1 dx^2] + 2[dx^3 dx^4] + \dots$$

Hence, in this special coordinate system, the matrix whose element in the α -th row and β -th column is $a_{\alpha\beta}$ has the form

$$(2) \quad \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} + \dots + \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} + \begin{pmatrix} & \\ & \end{pmatrix}$$

* Received December 1, 1946.

¹ H. C. Lee, "A kind of even-dimensional differential geometry and its application to exterior calculus," *American Journal of Mathematics*, vol. 65 (1943), pp. 433-438; "On even-dimensional skew-metric spaces and their groups of transformations," *ibid.*, vol. 67 (1945), pp. 321-328. These will be referred to as (I) and (II).

² The indices $\alpha, \beta, \gamma, \rho, \sigma, \tau$ run through the range $1, \dots, m$.

³ Cf. (I), p. 434. The reasoning applies whether $a_{\alpha\beta}$ is singular or nonsingular.

where the 2×2 matrix $\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ is repeated n times, and the last direct summand $\begin{pmatrix} & \\ & \end{pmatrix}$ denotes the zero matrix of order $m - 2n$. By rearrangement of rows and columns we have

THEOREM 1. *For a flat space L_m of rank $2n$, there exists a "canonical" coordinate system in which the fundamental tensor $a_{\alpha\beta}$ has constant components given by a matrix of the form*

$$(3) \quad \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix} \dot{+} \begin{pmatrix} & \\ & \end{pmatrix},$$

where I and 0 denote the unit and zero matrices of order n respectively, and $\begin{pmatrix} & \\ & \end{pmatrix}$ is the zero matrix of order $m - 2n$.

If m is even and if L_m is of rank m , the covariant skewsymmetric tensor $a_{\alpha\beta}$ being nonsingular has an inverse $a^{\alpha\beta}$ which is a contravariant skewsymmetric tensor. But if $a_{\alpha\beta}$ is singular (certainly so if m is odd), $a^{\alpha\beta}$ does not exist. We are therefore led to introduce skew-metric spaces of another kind as follows.

1. **The space L^m .** Let there be given a skewsymmetric contravariant tensor $a^{\alpha\beta}(x)$, of rank $2n$ say. An m -dimensional space equipped with such a fundamental tensor is called an L^m of rank $2n$. The skewsymmetric quantity

$$(4) \quad K^{\alpha\beta\gamma} = a^{\alpha\rho} \frac{\partial a^{\beta\gamma}}{\partial x^\rho} + a^{\beta\rho} \frac{\partial a^{\gamma\alpha}}{\partial x^\rho} + a^{\gamma\rho} \frac{\partial a^{\alpha\beta}}{\partial x^\rho}$$

is a contravariant tensor, which we call the *curvature tensor* of L^m .

When (and only when) m is even and $2n = m$, the space L^m , of rank m , is said to be *nonsingular*. In this case $a^{\alpha\beta}$ has an inverse $a_{\alpha\beta}$, and the space L^m is at the same time a space L_m . The two tensors (1) and (4) are now related to each other by the equations⁴

$$K^{\alpha\beta\gamma} = a^{\alpha\rho} a^{\beta\sigma} a^{\gamma\tau} K_{\rho\sigma\tau}, \quad K_{\alpha\beta\gamma} = a_{\alpha\rho} a_{\beta\sigma} a_{\gamma\tau} K^{\rho\sigma\tau}.$$

In any case, a space L^m is said to be *flat* if its curvature tensor $K^{\alpha\beta\gamma}$ vanishes:

$$(5) \quad a^{\alpha\rho} \frac{\partial a^{\beta\gamma}}{\partial x^\rho} + a^{\beta\rho} \frac{\partial a^{\gamma\alpha}}{\partial x^\rho} + a^{\gamma\rho} \frac{\partial a^{\alpha\beta}}{\partial x^\rho} = 0.$$

The question arises whether there exist special coordinate systems for a flat L^m , in which the fundamental tensor $a^{\alpha\beta}$ has constant components. The

⁴ See (II), p. 324.

answer is immediate in case m is even and L^m is nonsingular: for then L^m is also an L_m , the fundamental tensors $a^{a\beta}$ and $a_{a\beta}$ of L^m and L_m being inverse to each other, and by Theorem 1 there exists a canonical coordinate system of L_m in which $a_{a\beta}$ is given by

$$(6) \quad \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix},$$

whence, in this coordinate system, $a^{a\beta}$ is

$$(7) \quad \begin{pmatrix} 0 & -I \\ I & 0 \end{pmatrix}.$$

However, this proof is invalid in case L^m is singular. We shall show that for any flat L^m , the answer to the above question is in the affirmative.

2. Some properties of flat spaces. In passing we note certain formulas which will be of use later on. Consider any space L^m , not necessarily flat. For any two functions $f(x)$, $g(x)$ of the coordinates we form the (generalised) *Poisson expression*

$$(8) \quad (f, g) = a^{a\beta} \frac{\partial f}{\partial x^a} \frac{\partial g}{\partial x^\beta},$$

which is skewsymmetric in f and g :

$$(f, g) = -(g, f).$$

For any three functions $f(x)$, $g(x)$, $h(x)$, it is easy to verify the relation

$$(9) \quad (f, (g, h)) + (g, (h, f)) + (h, (f, g)) = K^{a\beta\gamma} \frac{\partial f}{\partial x^a} \frac{\partial g}{\partial x^\beta} \frac{\partial h}{\partial x^\gamma}.$$

Hence, when L^m is flat, the Jacobi identity

$$(10) \quad (f, (g, h)) + (g, (h, f)) + (h, (f, g)) = 0$$

is valid. If we define the operators

$$(11) \quad \Delta^a = a^{a\beta} \frac{\partial}{\partial x^\beta},$$

called the *fundamental operators* of the space L^m , we find

$$\begin{aligned} \Delta^a \Delta^\beta - \Delta^\beta \Delta^a &= a^{a\rho} \frac{\partial}{\partial x^\rho} \left(a^{\beta\sigma} \frac{\partial}{\partial x^\sigma} \right) - a^{\beta\sigma} \frac{\partial}{\partial x^\sigma} \left(a^{a\rho} \frac{\partial}{\partial x^\rho} \right) \\ &= \left(a^{a\rho} \frac{\partial a^{\beta\gamma}}{\partial x^\rho} + a^{\beta\rho} \frac{\partial a^{\gamma a}}{\partial x^\rho} \right) \frac{\partial}{\partial x^\gamma} \\ &= \left(K^{a\beta\gamma} - a^{\gamma\rho} \frac{\partial a^{a\beta}}{\partial x^\rho} \right) \frac{\partial}{\partial x^\gamma}, \end{aligned}$$

and so we have the formula

$$(12) \quad \Delta^a \Delta^\beta - \Delta^\beta \Delta^a = \frac{\partial a^{a\beta}}{\partial x^\gamma} \Delta^\gamma + K^{a\beta\gamma} \frac{\partial}{\partial x^\gamma}.$$

In particular, when L^m is flat, we have the relation

$$(13) \quad \Delta^a \Delta^\beta - \Delta^\beta \Delta^a = \frac{\partial a^{a\beta}}{\partial x^\gamma} \Delta^\gamma,$$

which states that the commutators of the operators Δ^a are linear combinations of these operators themselves. Hence the m fundamental operators $\Delta^1, \dots, \Delta^m$ of a flat space L^m constitute a "complete system."

3. Canonical coordinate systems of a flat space L^m . We are now in position to show that for any flat space L^m we can determine a coordinate system in which the components $a^{a\beta}$ of the fundamental tensor are all constants.

If all the $a^{a\beta}$ vanish, there is no more to prove.

In the other case, take any $a^{\rho\sigma} \neq 0$ where ρ and σ are fixed, $\rho \neq \sigma$. If neither ρ nor σ is 1, we interchange the roles of the coordinates x^ρ and x^1 . In other words, we can always choose the *new coordinate* x^1 (here by a mere renumbering of the coordinates) such that $a^{1\beta} \neq 0$ for at least a value of β ($\neq 1$).

Then we set up the equation

$$a^{1\beta} \frac{\partial f}{\partial x^\beta} = -1,$$

whose solutions do not depend on x^1 alone (since $a^{11} = 0$). Choosing any solution of this equation as the *new coordinate* x^2 , we have $a^{1\beta} \frac{\partial x^2}{\partial x^\beta} = -1$ so that

$$a^{12} = -1.$$

Consider now the two equations

$$a^{1\beta} \frac{\partial f}{\partial x^\beta} = 0, \quad a^{2\beta} \frac{\partial f}{\partial x^\beta} = 0$$

which are independent since $a^{11} = 0$, $a^{12} = -1$, $a^{21} = 1$, $a^{22} = 0$; they can be solved for the two derivatives $\frac{\partial f}{\partial x^2}$ and $\frac{\partial f}{\partial x^1}$ as linear combinations of the remaining derivatives. The corresponding operators are Δ^1 and Δ^2 , and their commutator is by (13)

$$\Delta^1 \Delta^2 - \Delta^2 \Delta^1 = \frac{\partial(-1)}{\partial x^1} \Delta^1 = 0.$$

Hence the two equations in question constitute a complete system, which, therefore, has $m - 2$ independent solutions. Let f^p ($p = 3, \dots, m$) be any set of such solutions. These functions being independent, the matrix

$$\left(\frac{\partial f^p}{\partial x^1}, \frac{\partial f^p}{\partial x^2}, \dots, \frac{\partial f^p}{\partial x^m} \right)$$

of $m - 2$ rows and m columns is of rank $m - 2$. Since the first two columns of this matrix are linear combinations of the remaining $m - 2$ columns, these $m - 2$ columns must have a nonzero determinant, which is the functional determinant $\frac{\partial(f^3, \dots, f^m)}{\partial(x^3, \dots, x^m)}$. Then

$$\frac{\partial(x^1, x^2, f^3, \dots, f^m)}{\partial(x^1, x^2, x^3, \dots, x^m)} = \frac{\partial(f^3, \dots, f^m)}{\partial(x^3, \dots, x^m)} \neq 0,$$

and therefore the m functions $x^1, x^2, f^3, \dots, f^m$ are independent. Hence it is justified to choose f^3, \dots, f^m as the *new coordinates* x^3, \dots, x^m . If in the conditions

$$a^{1\beta} \frac{\partial f^p}{\partial x^\beta} = 0, \quad a^{2\beta} \frac{\partial f^p}{\partial x^\beta} = 0 \quad (p = 3, \dots, m)$$

we set $f^p = x^p$, we have

$$a^{1p} = 0, \quad a^{2p} = 0 \quad (p = 3, \dots, m).$$

We recall then the identities (5). Giving the free indices α, β, γ various values we have that the components a^{pq} ($p, q = 3, \dots, m$) are functions of the coordinates x^3, \dots, x^m only, independent of x^1 and x^2 , and that they satisfy the identities

$$a^{ps} \frac{\partial a^{qr}}{\partial x^s} + a^{qs} \frac{\partial a^{rp}}{\partial x^s} + a^{rs} \frac{\partial a^{pq}}{\partial x^s} = 0 \quad (p, q, r, s = 3, \dots, m).$$

Consider a space of $m - 2$ dimensions with the coordinates x^3, \dots, x^m . A transformation of coordinates in this space is effected by $m - 2$ independent functions

$$\bar{x}^3 = \phi^3(x^3, \dots, x^m), \dots, \bar{x}^m = \phi^m(x^3, \dots, x^m).$$

It is easily seen that the components a^{pq} ($p, q = 3, \dots, m$) constitute a contravariant tensor in this space. In fact, if in the known transformation law

$$\bar{a}^{\rho\sigma} = a^{\alpha\beta} \frac{\partial \bar{x}^\rho}{\partial x^\alpha} \frac{\partial \bar{x}^\sigma}{\partial x^\beta}$$

of $a^{\alpha\beta}$ we restrict the free indices ρ and σ to the range $3, \dots, m$, we have

$$\bar{a}^{rs} = a^{pq} \frac{\partial \bar{x}^r}{\partial x^p} \frac{\partial \bar{x}^s}{\partial x^q} \quad (p, q, r, s = 3, \dots, m).$$

This proves the tensor character of a^{pq} in the space considered. This space, equipped with the skewsymmetric contravariant tensor a^{pq} , is an L^{m-2} . The curvature tensor of this L^{m-2} vanishes because of the identities satisfied by a^{pq} , whence the space L^{m-2} is flat.

For this flat L^{m-2} , the whole of the above proof may be repeated, with the result that either all $a^{pq} = 0$ or we can determine a new coordinate system in L^{m-2} for which

$$a^{34} = -1, \quad a^{3t} = a^{4t} = 0 \quad (t = 5, \dots, m).$$

Continuing this method of proof we finally arrive at a coordinate system in which the components of $a^{\alpha\beta}$ are given by the matrix

$$(14) \quad \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} + \dots + \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} + \begin{pmatrix} & \\ & \end{pmatrix}.$$

If $a^{\alpha\beta}$ is of rank $2n$, the 2-rowed matrix $\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ occurs n times, and the last term $\begin{pmatrix} & \\ & \end{pmatrix}$ is the zero matrix of order $m - 2n$. By a permutation of the coordinates we have the following theorem, analogous to Theorem 1:

THEOREM 2. *For a flat space L^m of rank $2n$, there exists a canonical coordinate system in which the fundamental tensor $a^{\alpha\beta}$ is given by*

$$(15) \quad \begin{pmatrix} 0 & -I \\ I & 0 \end{pmatrix} + \begin{pmatrix} & \\ & \end{pmatrix},$$

where I and 0 denote the unit and zero matrices of order n respectively, and $\begin{pmatrix} & \\ & \end{pmatrix}$ is the zero matrix of order $m - 2n$.

We note that in the nonsingular case, we have two different proofs of the same results (6) and (7).

4. Equivalence of flat spaces. Two spaces L^m (or two spaces L_m) are said to be *equivalent* if the fundamental tensor of one can be transformed into the fundamental tensor of the other by a change of coordinates. In consequence of Theorems 1 and 2, we have

THEOREM 3. *Two flat spaces L^m (or two flat spaces L_m) are equivalent if and only if they are of the same rank.*

5. Function groups. Consider a space L^m . Let there be r ($\leq m$)

Let us examine formula (17). Its left-hand member is the Poisson expression of f and g in L^m according to definition (8), and its right-hand member is, according to a similar definition, the Poisson expression of f and g in L^r . Hence

THEOREM 4. *For any two functions f and g in the group space L^r (which are also functions in the space L^m), the expression (f, g) has a double meaning, whether relative to the space L^m or to the group space L^r .*

This double meaning of the formation (f, g) has important consequences. Let f, g, h be any three functions in L^r , and consider the expression

$$(f, (g, h)) + (g, (h, f)) + (h, (f, g)).$$

Regarding it as relative to L^m , we have formula (9); and if we refer it to L^r we have the similar relation

$$(21) \quad (f, (g, h)) + (g, (h, f)) + (h, (f, g)) = R^{\lambda\mu\nu} \frac{\partial f}{\partial y^\lambda} \frac{\partial g}{\partial y^\mu} \frac{\partial h}{\partial y^\nu}.$$

Hence, comparing this with (9), we have

$$(22) \quad R^{\lambda\mu\nu} \frac{\partial f}{\partial y^\lambda} \frac{\partial g}{\partial y^\mu} \frac{\partial h}{\partial y^\nu} = K^{\alpha\beta\gamma} \frac{\partial f}{\partial x^\alpha} \frac{\partial g}{\partial x^\beta} \frac{\partial h}{\partial x^\gamma}.$$

If we set f, g, h in (22) equal to y^λ, y^μ, y^ν respectively, we find that the curvature tensor of the group space L^r is given by

$$(23) \quad R^{\lambda\mu\nu} = K^{\alpha\beta\gamma} \frac{\partial y^\lambda}{\partial x^\alpha} \frac{\partial y^\mu}{\partial x^\beta} \frac{\partial y^\nu}{\partial x^\gamma}.$$

In particular, when L^m is flat: $K^{\alpha\beta\gamma} = 0$, we have also $R^{\lambda\mu\nu} = 0$. Hence

THEOREM 5. *The group space of a function group of order r in a flat space L^m is a flat space L^r .*

7. Function groups in flat spaces. Consider a function group of order r in a flat space L^m . The group space L^r of the group is flat by Theorem 5. Hence, by Theorem 2, there exists a canonical coordinate system for L^r in which the fundamental tensor $b^{\lambda\mu}$ of L^r (defined by (16)) reduces to a form like (15). Consequently,

THEOREM 6. *For a function group of order r in a flat space L^m , there exists a canonical basis $\{y^1, \dots, y^r\}$ such that the Poisson expressions (y^λ, y^μ) are given by a matrix of the form*

$$(24) \quad \begin{pmatrix} 0 & -I \\ I & 0 \end{pmatrix} + \begin{pmatrix} & \\ & \end{pmatrix}$$

where I and 0 denote the unit and zero matrices of order s and (\quad) is the zero matrix of order $r - 2s$.^o

The rank of the r -rowed skewsymmetric matrix (y^λ, y^μ) is $2s$, which we call the *rank* of the group. This is also the rank of the group space L^r .

Take any p ($< r$) of the functions y^1, \dots, y^r of a canonical basis. The Poisson parenthesis of these functions being constants ($= 1, 0$), they can be regarded as functions of the p functions themselves. Hence these p functions, being independent, constitute a basis of a function group of order p , which is a subgroup of the given group. We have then

THEOREM 7. *A function group in a flat space L^m contains subgroups of every lower order.*

The set of all functions of the coordinates x of L^m is evidently a function group of order m . Hence

THEOREM 8. *In a flat space L^m there exist function groups of every order $r = 1, \dots, m$.*

8. Induced operators. Consider a general space L^m and the group space L^r of a function group, as in 5 and 6. With (11) we define the induced operators

$$(25) \quad \delta^\lambda = \frac{\partial y^\lambda}{\partial x^a} \Delta^a.$$

It is required to find the commutators $\delta^\lambda \delta^\mu - \delta^\mu \delta^\lambda$ of these operators. We proceed as follows.

We first note that for a function g of the coordinates x we have in view of (25) and (11)

$$(26) \quad \delta^\lambda g = (y^\lambda, g).$$

Hence

$$\begin{aligned} \delta^\lambda \delta^\mu g - \delta^\mu \delta^\lambda g &= \delta^\lambda (y^\mu, g) - \delta^\mu (y^\lambda, g) \\ &= (y^\lambda, (y^\mu, g)) - (y^\mu, (y^\lambda, g)) \\ &= - (g, (y^\lambda, y^\mu)) + K^{\alpha\beta\gamma} \frac{\partial y^\lambda}{\partial x^\alpha} \frac{\partial y^\mu}{\partial x^\beta} \frac{\partial g}{\partial x^\gamma} \end{aligned}$$

by (9), and since

^o This is a generalisation of a classical theorem. See L. P. Eisenhart, *Continuous groups of transformations*, p. 285.

$$-(g, (y^\lambda, y^\mu)) = (b^{\lambda\mu}, g) = \frac{\partial b^{\lambda\mu}}{\partial y^\nu} (y^\nu, g) = \frac{\partial b^{\lambda\mu}}{\partial y^\nu} \delta^\nu g$$

we obtain the formula

$$(27) \quad \delta^\lambda \delta^\mu - \delta^\mu \delta^\lambda = \frac{\partial b^{\lambda\mu}}{\partial y^\nu} \delta^\nu + K^{\alpha\beta\gamma} \frac{\partial y^\lambda}{\partial x^\alpha} \frac{\partial y^\mu}{\partial x^\beta} \frac{\partial}{\partial x^\gamma}.$$

If in particular the operand g (as function of the x) is a function of the functions y , we may write

$$\frac{\partial}{\partial x^\gamma} = \frac{\partial y^\nu}{\partial x^\gamma} \frac{\partial}{\partial y^\nu}$$

in (27) and on account of (23) we have

$$(28) \quad \delta^\lambda \delta^\mu - \delta^\mu \delta^\lambda = \frac{\partial b^{\lambda\mu}}{\partial y^\nu} \delta^\nu + R^{\lambda\mu\nu} \frac{\partial}{\partial y^\nu}.$$

It should be remarked that in this case the right-hand side of (26) may be written $(y^\lambda, y^\mu) \frac{\partial g}{\partial y^\nu}$, so that we have

$$(29) \quad \delta^\lambda = b^{\lambda\mu} \frac{\partial}{\partial y^\mu}.$$

Hence, *operating on functions of the y 's, the δ 's are the fundamental operators of the group space L^ν in the same way as the Δ 's are fundamental operators of the space L^m .* With this in view, and recalling (12) in connection with (11), we could have written down formula (28) without proof.

In particular, *when L^m is flat (then L^ν is also flat), (27) or (28) reduces to*

$$(30) \quad \delta^\lambda \delta^\mu - \delta^\mu \delta^\lambda = \frac{\partial b^{\lambda\mu}}{\partial y^\nu},$$

and so *the operators $\delta^1, \dots, \delta^r$ constitutes a complete system* (whether for the space L^m or for the group space L^r).

9. The reciprocal group and the centrum of a function group in flat space. Suppose that the space L^m is flat. Consider then the complete system of equations

$$(31) \quad \delta^\lambda g = 0 \quad (\lambda = 1, \dots, r),$$

which are not necessarily independent. For every solution of this system we have $(y^\lambda, g) = 0$ by (26), whence $(f, g) = 0$ for all functions $f(y)$ of the function group. We say that g is *involution* or *commutative* with f .

If g is regarded as a function of the variables x^1, \dots, x^m , the number k of independent equations in the system (31) is equal to the rank of the

product matrix $\frac{\partial y^\lambda}{\partial x^\alpha} a^{\alpha\beta}$ (see (25)); thus $k \leq \min(r, 2n)$ where $2n$ is the rank of $a^{\alpha\beta}$. In particular, if L^m is nonsingular (i. e., $m = 2n$), then $k = r$. The system (31) has $m - k$ independent solutions, $g^p(x)$ ($p = 1, \dots, m - k$). For any two of these solutions we have

$$\begin{aligned} \delta^\lambda(g^p, g^q) &= (y^\lambda, (g^p, g^q)) \\ &= -(g^p, (y^\lambda, g^q)) + (g^q, (y^\lambda, g^p)) \\ &= -(g^p, \delta^\lambda g^q) + (g^q, \delta^\lambda g^p) \\ &= 0, \end{aligned}$$

so that (g^p, g^q) is again a solution of (31) and is therefore a function of the functions g^1, \dots, g^{m-k} . It follows that these functions constitute the basis of a function group of order $m - k$. Hence

THEOREM 9. *A function group in a flat space L^m determines another function group, the "reciprocal group," consisting of functions in involution with every function of the given group.*

If g is regarded as a function of the functions y^1, \dots, y^r the number of independent equations in the system (31) is (see (29)) equal to $2s$, the rank of $b^{\lambda\mu}$. This system has $r - 2s$ independent solutions, $g^p(y)$ ($p = 1, \dots, r - 2s$), which are commutative with every function of the given function group. The functions g^1, \dots, g^{r-2s} being themselves functions of the group, they constitute the basis of a subgroup. Hence

THEOREM 10. *The "centrum" of a function group of order r and rank $2s$ in a flat space L^m , consisting of all those functions of the group which are commutative with every function of the group, is a subgroup of order $r - 2s$. This subgroup is the intersection of the given group and its reciprocal group.*

STUDY OF A SURFACE BY MEANS OF CERTAIN ASSOCIATE RULED SURFACES IN AFFINE SPACE.*

By GEORGE WU.

In this paper we shall deduce some results concerning the theory of surfaces in affine space. Some of these results may be regarded as analogues of the projective theory of surfaces immersed in ordinary space. Others seem to be properties of surfaces immersed in affine space. In what follows we shall use the same notations as in Blaschke's *Vorlesungen über Differentialgeometrie* II.

First, we define the canonical quadric Q at an ordinary point P of a surface σ in affine space by means of the Bompiani-Kloboucek's asymptotic osculating quadrics. Using this quadric Q we give a simple geometrical interpretation of the Pick invariant J and the Gaussian curvature S of the fundamental quadric form $\phi = 2Fdudv$ of σ . In particular if $S = 0$, Q is a paraboloid, and conversely.

Next we study the Moutard quadrics $Q_n^{(u)}$ and $Q_n^{(v)}$ belonging to the tangent t_n and the asymptotic ruled surfaces $R^{(u)}$ and $R^{(v)}$ generated by the asymptotic u - and v -tangents along the v - and u -curves respectively. Using these two quadrics $Q_n^{(u)}$ and $Q_n^{(v)}$, we give a characteristic property of the tangents of Segre. We have shown that the loci of the diameters $\bar{d}_n^{(u)}$ and $\bar{d}_n^{(v)}$ of $Q_n^{(u)}$ and $Q_n^{(v)}$ are two quadric cones $\Gamma_2^{(u)}$ and $\Gamma_2^{(v)}$ passing respectively through the asymptotic u - and v -tangents and having the affine surface normal as a common generator. The cones $\Gamma_2^{(u)}$ and $\Gamma_2^{(v)}$ intersect the quadric of Lie in the asymptotic tangents and in a pair of non-composite twisted cubics $C_3^{(u)}$ and $C_3^{(v)}$. The tangent plane π of σ at P cuts the tangent surface of $C_3^{(u)}$ and $C_3^{(v)}$ in a pair of parabolas, mutually intersecting, besides the point P , and at three other points lying on the tangents of Segre. This property of the Segre tangents is similar to properties demonstrated by Čech¹ and Su.²

Analogous to a theorem of Transon we prove that the loci of the affine

* Received February 27, 1947.

¹ E. Čech, "L'intorno di una superficie considerate dal punto di vista proiettivo," *Annali di Matematica*, vol. (3) 31 (1922), p. 205.

² B. Su, "The quadric of Moutard I," *Tohoku Mathematical Journal*, vol. 33 (1931), p. 35.

normals of the sections of $R^{(u)}$ and $R^{(v)}$ made by any plane π_{np} as it rotates about a non-asymptotic tangent t_n are two planes $\pi_n^{(u)}$ and $\pi_n^{(v)}$. As t_n varies, $\pi_n^{(u)}$ and $\pi_n^{(v)}$ envelop two cones which coincide with $\Gamma_n^{(u)}$, $\Gamma_n^{(v)}$. Moreover, the planes $\pi_n^{(u)}$, $\pi_n^{(v)}$ and the Transon plane T are concurrent in the conjugate tangent of t_n . The envelopes of the harmonic conjugate planes of π and T with respect to $\pi_n^{(u)}$ and $\pi_n^{(v)}$ are found. The planes $\pi_n^{(u)}$ and $\pi_n^{(v)}$ and those corresponding to the conjugate tangent t_n are studied in detail.

1. **Analytic basis.** If we assume the asymptotic curves to be parametric then the coordinates \mathfrak{x} of any non-parabolic point P of a non-ruled analytic surface σ satisfy the following system of completely integrable equations:

$$(1) \quad \begin{cases} \mathfrak{x}_{uu} = (F_u/F)\mathfrak{x}_u + (A/F)\mathfrak{x}_v, \\ \mathfrak{x}_{uv} = F\mathfrak{y}, \\ \mathfrak{x}_{vv} = (D/F)\mathfrak{x}_u + (F_v/F)\mathfrak{x}_v; \end{cases}$$

$$(2) \quad \begin{cases} \mathfrak{y}_u = -H\mathfrak{x}_u + (A_v/F)\mathfrak{x}_v, \\ \mathfrak{y}_v = + (D_u/F)\mathfrak{x}_u - H\mathfrak{x}_v. \end{cases}$$

From (1) and (2) we have

$$(3) \quad \begin{cases} \mathfrak{x}_{uuu} = (F_{uu}/F)\mathfrak{x}_u + (A_u/F)A\mathfrak{y}, \\ \mathfrak{x}_{uuv} = -FH\mathfrak{x}_u + (A_v/F)\mathfrak{x}_v + F_u\mathfrak{y}, \\ \mathfrak{x}_{uvv} = (D_u/F)\mathfrak{x}_u - 2FH\mathfrak{y} + F_v\mathfrak{y}, \\ \mathfrak{x}_{vvv} = (D_v/F)\mathfrak{x}_u + (F_{vv}/F)\mathfrak{x}_v + D\mathfrak{y}; \\ \mathfrak{x}_{uuuu} = (*)\mathfrak{x}_u + (*)\mathfrak{x}_v + 2A_u\mathfrak{y}, \\ \mathfrak{x}_{uuuv} = (*)\mathfrak{x}_u + (*)\mathfrak{x}_v + (A_v + F_{uu})\mathfrak{y}, \end{cases}$$

where $(*)$ denotes terms we shall not need. Let us denote the coordinates of any point (\mathfrak{z}) in space by

$$\mathfrak{z} = \mathfrak{x} + x\mathfrak{x}_u + y\mathfrak{x}_v + z\mathfrak{y}.$$

Then we have

$$\begin{cases} x = + (1/F)(\mathfrak{z} - \mathfrak{x}, \mathfrak{x}_v, \mathfrak{y}), \\ y = - (1/F)(\mathfrak{z} - \mathfrak{x}, \mathfrak{x}_u, \mathfrak{y}), \\ z = + (1/F)(\mathfrak{z} - \mathfrak{x}, \mathfrak{x}_u, \mathfrak{x}_v). \end{cases}$$

2. **The canonical quadric.** Lane³ has defined the canonical quadric by using Bompiani-Kloubouček's asymptotic osculating quadrics. In an

³ E. P. Lane, *Projective Differential Geometry of Curves and Surfaces*, University of Chicago Press, p. 119. Ex. 38.

analogous manner we can obtain the canonical quadric at a point of a surface in affine space. The equations of the asymptotic osculating quadrics⁴ corresponding to the curve C_λ through P immersed in the one-parameter family of curves,

$$dv - \lambda du = 0,$$

in affine space are

$$Q^{(v)} \quad H z^2 - 2z + 2Fxy + 2(D/F)\lambda^2xz + 2(D/F)\lambda yz \\ - (A/F^2)\{A/F + \lambda(A_u/F) + \lambda^2(\partial/\partial v)\log(A^2/F) - \lambda'\}z^2 = 0$$

and

$$Q^{(u)} \quad H z^2 - 2z + 2Fxy + 2(D/F)\lambda^2xz + 2(D/F)\lambda yz \\ - (D/F^2)\{(D/F)\lambda^3 + \lambda^2(D_v/F) + \lambda(\partial/\partial u)\log(D^2/F) + \lambda'\}z^2 = 0$$

where λ' is the total differential of λ with respect to u .

The equation of the osculating plane of C_λ at the point P is

$$2F\lambda(\lambda x - y) + \{A/F - (F_u/F)\lambda + (F_v/F)\lambda^2 - (D/F)\lambda^3 + \lambda'\}z = 0.$$

This plane intersects the quadric $Q^{(u)}$ in a conic. The locus of this conic as C_λ varies but remains touching a fixed tangent t_λ at P is the quadric whose equation is

$$(4) \quad \bar{Q}^{(u)}: \lambda^3(Sz^2 - 2z + 2Fxy) - 4(A/F)\lambda^2xz + 4(A/F)\lambda yz \\ - (A/F^2)\{2A/F + \lambda(\partial/\partial u)\log(A/F) + 2\lambda^2(\partial/\partial v)\log A\}z^2 = 0.$$

On replacing the quadric $Q^{(u)}$ by $Q^{(v)}$, we find another quadric:

$$(5) \quad \bar{Q}^{(v)}: Sz^2 - 2z + 2Fxy + 4(D/F)\lambda^2xz - 4(D/F)\lambda yz \\ - (D/F^2)\{(2D/F)\lambda^3 + \lambda^2(\partial/\partial u)\log D/F + 2\lambda(\partial/\partial u)\log D\}z^2 = 0.$$

When $\lambda \rightarrow \infty$ in (4) and $\lambda \rightarrow \infty$ in (5), both $\bar{Q}^{(u)}$ and $\bar{Q}^{(v)}$ approach the same quadric Q :

$$2Fxy - 2z + Sz^2 = 0$$

where

$$S = J + H = -\frac{1}{F} \frac{\partial^2}{\partial u \partial v} \log F,$$

S being the Gaussian curvature of $\phi = 2Fdu dv$. It is well known that the quadric Q is the canonical quadric of σ at P . Hence we have the

⁴ B. Su, "A note on the affine differential geometry of a surface," *Japanese Journal of Mathematics*, vol. 9 (1932), p. 235.

THEOREM. If $S = 0$, Q is a paraboloid and conversely.

The center (\mathfrak{z}) of the quadric Q is

$$(6) \quad \mathfrak{z} = \mathfrak{x} + (1/S)\mathfrak{y}.$$

That is to say the reciprocal of the affine distance p of the center (\mathfrak{z}) of Q from the point P is the Gaussian curvature S . Let us denote the affine distance of the center of the quadric of Lie F_2 whose equation is

$$(7) \quad 2Fxy - 2z + Hz^2 = 0$$

from P by p' ; then $p' = 1/H$. The Pick invariant J of σ is equal to $1/p - 1/p'$. It is easy to see that a necessary and sufficient condition for $Q \equiv F_2$ is $J = 0$. That is to say: $Q \equiv F_2$ if and only if σ is a ruled surface.

From (6) we have

$$\begin{aligned} \mathfrak{z}_u &= (J/S)\mathfrak{x}_u + (A_v/F^2S)\mathfrak{x}_v + (1/S)u\mathfrak{y}, \\ \mathfrak{z}_v &= (D_u/F^2S)\mathfrak{x}_u + (J/S)\mathfrak{x}_v + (1/S)v\mathfrak{y}. \end{aligned}$$

Hence a necessary and sufficient condition for (\mathfrak{z}) to be a fixed point in space is that σ be a ruled affine sphere ($J = A_v = D_u = 0$).

3. The Moutard quadrics of $R^{(u)}$ and $R^{(v)}$. Let us consider the asymptotic ruled surface $R^{(u)}$ [$R^{(v)}$] generated by the asymptotic u -tangent [v -tangent] along the v -curve [u -curve]. It is evident that the point

$$\bar{\mathfrak{z}} = \mathfrak{x} + \mathfrak{x}_v$$

lies on the u -tangent of \mathfrak{x} . Hence we may take the v -curve and the curve generated by $\bar{\mathfrak{z}}$ as the director curves of $R^{(u)}$. Thus we have

$$\begin{aligned} \mathfrak{x}(u + du, 0) &= \mathfrak{x} + \mathfrak{x}_u du + \mathfrak{x}_{uu}(du^2/2!) + \mathfrak{x}_{uuu}(du^3/3!) + \mathfrak{x}_{uuuu}(du^4/4!) + (5), \\ \mathfrak{z}(u + du, 0) &= \mathfrak{x} + \mathfrak{x}_v + (\mathfrak{x}_u + \mathfrak{x}_{uv}du + (\mathfrak{x}_{uu} + \mathfrak{x}_{uuv})(du^2/2!) \\ &\quad + (\mathfrak{x}_{uuu} + \mathfrak{x}_{uuuv})(du^3/3!) + (4). \end{aligned}$$

If we write

$$\begin{aligned} \mathfrak{x}(u + du, 0) &= \mathfrak{x} + \xi^{(0)}\mathfrak{x}_u + \eta^{(0)}\mathfrak{x}_v + \zeta^{(0)}\mathfrak{y}, \\ \bar{\mathfrak{z}}(u + du, 0) &= \mathfrak{x} + \xi^{(1)}\mathfrak{x}_u + \eta^{(1)}\mathfrak{x}_v + \zeta^{(1)}\mathfrak{y} \end{aligned}$$

then in view of (3) we find

$$\begin{cases} \xi^{(0)} = du + \frac{F_u}{F} \frac{du^2}{2!} + \frac{F_{uu}}{F} \frac{du^3}{3!} + (4), \\ \eta^{(0)} = \frac{A}{F} \frac{du^2}{2!} + \frac{A_u}{F} \frac{du^3}{3!} + (4), \\ \zeta^{(0)} = A(du^3/3!) + 2A_u(du^4/4!) + (5); \\ \begin{cases} \xi^{(1)} = du + [(F_u/F) - FH](du^2/2!) + (3), \\ \eta^{(1)} = 1 + [A/F + A_v/F](du^2/2!) + (3), \\ \zeta^{(1)} = Fdu + F_u(du^2/2!) + [A + A_v + F_{uu}](du^3/3!) + (4). \end{cases} \end{cases}$$

The coordinates of any point on $R^{(u)}$ may be expressed in the form

$$(8) \quad x = \frac{\xi^{(0)} + p\xi^{(1)}}{1+p}, \quad y = \frac{\eta^{(0)} + p\eta^{(1)}}{1+p}, \quad z = \frac{\zeta^{(0)} + p\zeta^{(1)}}{1+p}$$

where p is to be regarded as a parameter and x, y, z as current coordinates.

The equation of one of the planes π_{np} passing through the non-asymptotic tangent t_n :

$$z = y - nx = 0$$

is of the form

$$(9) \quad z = \rho(y - nx)$$

where ρ is a parameter. We demand that the parameter p be such that the point (8) lies on the plane (9). Substituting (8) into (9) we get

$$(10) \quad p = ndu + \{2n^2 + nF_u/F - A/F + 2nF/\rho\}du^2/2! \\ + \{6n^3 + n^2[6(F_u/F) - 3FH] + n[F_{uu}/F - 6A/F - 3A_v/F] \\ - A_u/F - 2A/\rho + 6nF_u/\rho + 12n^2F/\rho + 6nF^2/\rho^2\}du^3/3! + (4),$$

and

$$(11) \quad (1+p)^{-1} = 1 - ndu + \{-nF_u/F + A/F - 2nF/\rho\}du^2/2! + (3).$$

Substituting (10) and (11) into (8) we get the expansion of x up to the third order, namely:

$$x = du + \frac{F_u}{F} \frac{du^2}{2!} + \{-3nFH + (F_{uu}/F)\}du^3/3! + (4);$$

and that of z up to fourth order namely:

$$z = nFdu^2 + \{6nF_u - 2A + 6n(F/\rho)\}du^3/3! \\ + \{n^2[-12F^2H] + n[8F_{uu} - 8A_v + 6(F^2_u/F)] - 2A_u - 8AF/\rho \\ + 36n(FF_u/\rho) + 24n(F^3/\rho)\}du^4/4! + (5).$$

It is not difficult to express z as power series in x , in the form

$$(12) \quad z = a_2x^2 + a_3x^3 + a_4x^4 + (5),$$

wherein we have placed

$$a_2 = nF,$$

$$a_3 = 1/6[-2A + 6n(F^2/\rho)],$$

$$a_4 = 1/24[12n^2FH - 8nA_v - 2A_u + 6A(F_u/F) - 8A(F/\rho) + 24n(F^2/\rho^2)].$$

Evidently (12) is the projection of the section of $R^{(u)}$ made by the plane π_{np} on the xz -plane. The osculating conic of (12) at P is

$$(13) \quad -x^2 + z(\alpha x + \beta z + \gamma) = 0$$

where α, β, γ have the values

$$\alpha = A/3nF^2 - 1/n\rho,$$

$$\beta = -\frac{1}{2} \frac{H}{nF} + A_v/3n^2F^3 + A_u/12n^3F^3 - AF_u/4n^3F^4$$

$$+ A^2/9n^4F^4 - A/3n^3F^2\rho,$$

$$\gamma = nF.$$

By eliminating ρ from (9) and (13) we get the equation of the Moutard quadric¹⁰ of $R^{(u)}$ belonging to the tangent t_n at P in the form

$$\begin{aligned} Q_n^{(u)} : & 36n^3F^2(z - Fxy) + 24n^2AF^2x^2 - 12nAF^2y^2 \\ & + \{4A^2 - 3Fn[A(\partial/\partial u) \log(F^3/A) - 4nA_v \\ & - 6n^2F(\partial^2/\partial u\partial v) \log F - 6n^2AD/F]\}z^2 = 0. \end{aligned}$$

Similarly we get the Moutard quadric¹⁰ of $R^{(v)}$ belonging to the tangent t_n at P in the form

$$\begin{aligned} Q_n^{(v)} : & 36F^2(z - Fxy) - 12n^2DF^2xz + 24nDF^2yz \\ & + \{4n^2D^2 - 3F[n^2D(\partial/\partial v) \log F^3/D - 4nDu \\ & - 6F(\partial^2/\partial u\partial v) \log F - 6AD/F]\}z^2 = 0. \end{aligned}$$

The residual conic K of the intersection of $Q_n^{(u)}$ and $Q_n^{(v)}$ lies in the plane

¹⁰ For the projective equivalence of these two quadrics see my paper, "Systems of quadrics associated with a point of a surface I, II," *Duke Mathematical Journal*, vol. 10 (1943), pp. 499-513, 515-530.

$$12n^2F^2(2A + n^3D)x - 12nF^2(A + 2n^3D)y \\ + \{4(A^2 - n^6D^2) - 3Fn[A(\partial/\partial u)\log F^3/A \\ - n^3D(\partial/\partial v)\log F^3/D + 4n^3D_u - 4A_v]\}z = 0.$$

A necessary and sufficient condition that K touch the tangent t_n at P is

$$A - n^3D = 0.$$

THEOREM. *The conic K is tangent to t_n if and only if t_n is a tangent of Segre.*

4. **Associate cones $\Gamma_2^{(u)}$ and $\Gamma_2^{(v)}$.** The diameter $d_n^{(u)}$ of $Q_n^{(u)}$ at P has the equations

$$\begin{cases} 2Az + 3nF^2y = 0, \\ Az - 3n^2F^2x = 0. \end{cases}$$

Similarly we have the diameter $d_n^{(v)}$ of $Q_n^{(v)}$ at P :

$$\begin{cases} 2nDz - 3F^2x = 0, \\ n^2Dz + 3F^2y = 0. \end{cases}$$

The loci of $d_n^{(u)}$ and $d_n^{(v)}$ as t_n varies in a pencil with the center P are the two cones

$$\Gamma_2^{(u)}: 3F^2y^2 + 4Axx = 0$$

and

$$\Gamma_2^{(v)}: 3F^2x^2 + 4Dyz = 0$$

which, we shall call the *associate cones* $\Gamma_2^{(u)}$ and $\Gamma_2^{(v)}$ respectively.

THEOREM. *The loci of $d_n^{(u)}$ and $d_n^{(v)}$ as t_n varies in a pencil with the center P are two quadric cones $\Gamma_2^{(u)}$ and $\Gamma_2^{(v)}$ passing respectively through the asymptotic u - and v -tangents with the affine surface normal as a common generator.*

A parametric representation of the quadric of Lie (7) is

$$(14) \quad x = \frac{2\lambda}{H\lambda\mu + 2F}, \quad y = \frac{2\mu}{H\lambda\mu + 2F}, \quad z = \frac{2\lambda\mu}{H\lambda\mu + 2F}.$$

Therefore, beside the asymptotic tangent $z = y = 0$, the intersection of (14) and $\Gamma_2^{(u)}$ is a twisted cubic $C_3^{(u)}$ whose equations are

$$(15) \quad x = \frac{3F^2\lambda}{3F^3 - 2AH\lambda^3}, \quad y = \frac{4A\lambda^2}{3F^3 - 2AH\lambda^3}, \quad z = \frac{4A\lambda^3}{3F^3 - 2AH\lambda^3}.$$

Similarly on replacing $\Gamma_2^{(u)}$ by $\Gamma_2^{(v)}$ we obtain another twisted cubic $C_3^{(v)}$ whose equations are

$$x = -\frac{4D\mu^2}{3F^3 - 2DH\mu^3}, \quad y = \frac{3F^2\mu}{3F^3 - 2DH\mu^3}, \quad z = -\frac{4D\mu^3}{3F^3 - 2DH\mu^3}.$$

Differentiating (15) with respect to λ we obtain

$$\frac{dx}{d\lambda} = \frac{3F^2(3F^3 + 4AH\lambda^3)}{(3F^3 - 2AH\lambda^3)^2}, \quad \frac{dy}{d\lambda} = \frac{-8A\lambda(3F^2 + AH\lambda)}{(3F^3 - 2AH\lambda^3)^2},$$

$$\frac{dz}{d\lambda} = \frac{-36AF^2\lambda^2}{(3F^3 - 2AH\lambda^3)^2}.$$

Hence the tangent surface of $C_3^{(u)}$ can be represented by

$$(16) \quad \begin{cases} x = \frac{3F^2\lambda}{3F^3 - 2AH\lambda^3} + 3F^2\rho[3F^3 + 4AH\lambda^3], \\ y = -\frac{4A\lambda^2}{3F^3 - 2AH\lambda^3} - 8A\lambda\rho[3F^3 + 4AH\lambda^3], \\ z = -\frac{4A\lambda^3}{3F^3 - 2AH\lambda^3} - 36AF^2\lambda^2\rho. \end{cases}$$

The section of (16) made by the tangent plane $z = 0$ is then

$$x = (2/3F)\lambda, \quad y = -(4A/9F^3)\lambda^2, \quad z = 0$$

or

$$(17) \quad x^2 + (F/A)y = 0, \quad z = 0$$

which has the asymptotic tangent $x = z = 0$ for its diameter and touches the asymptotic tangent $y = z = 0$. Replacing $C_3^{(u)}$ by $C_3^{(v)}$ we find another parabola

$$(18) \quad y^2 + (F/D)x = 0, \quad z = 0.$$

The two parabolas (17) and (18) intersect in the point P , and in three other points on the straight lines

$$z = Ax^3 - Dy^3 = 0.$$

Thus we have the

THEOREM. *The sections of the tangent surfaces of $C_3^{(u)}$ and $C_3^{(v)}$ made by the tangent plane at P are two parabolas, each of which touches one of the asymptotic tangents and has the other asymptotic tangent as its affine normal. They intersect in the point P , and in three other points lying on a tangent of Segre.*

The polar planes of the tangent t_n with respect to $\Gamma_2^{(u)}$ and $\Gamma_2^{(v)}$ determine a line

$$\begin{cases} 3nF^2y + 4Az = 0, \\ 2F^2x + 4nDz = 0. \end{cases}$$

By eliminating n from the above equations we obtain the equation of a cone of the second order

$$9F^4xy - 4ADz^2 = 0.$$

We call this cone *the associate cone* $\Gamma_2^{(uv)}$. Thus we have the

THEOREM. *The polar line of the tangent plane of σ at P with respect to the associate cone $\Gamma_2^{(uv)}$ is the affine surface normal.*

5. An analogue of a theorem of Transon. Transon⁵ has proved that all of the affine normals of the plane sections of σ made by π_{np} lie in a plane T . The equation of T is

$$3F^2(y + nx) - (A + Dn^3)z = 0.$$

This plane is the so called Transon Plane. Analogously we shall prove that the affine normals of the plane sections of $R^{(u)}$ [$R^{(v)}$] made by π_{np} lie also in a plane $\pi_n^{(u)}$ [$\pi_n^{(v)}$]. We call it *the associate plane* $\pi_n^{(u)}$ [$\pi_n^{(v)}$].

In order to prove the statement we adopt Salkowski's interpretation⁶ of the affine normal of a plane curve. He has stated that the affine normal at a point of a plane curve is the diameter of the parabola osculating the curve at this point. It is easily shown that the osculating parabola of the curve (12) at the point P is of the form

$$(19) \quad z = (rx + sz)^2, \quad y = 0$$

wherein we have placed

$$r = \sqrt{nF}, \quad s = 1/12r^3(-2A + 6nF^2/\rho).$$

Hence the affine normal of the curve (12) at P is

$$(20) \quad rx + (1/12r^3)(-2A + 6nF^2/\rho) = 0, \quad y = 0.$$

By eliminating ρ from (9) and (20) we obtain the equation of the associate plane $\pi_n^{(u)}$ in the form

⁵ A. Transon, "Recherches sur la théorie des lignes et des surfaces," *Journal des Mathématiques* (1), vol. 6 (1841), pp. 191-208.

⁶ E. Salkowski, *Affine Differentialgeometrie* (1935), S. 49-50.

$$(21) \quad 3F^2n(y + nx) - Az = 0.$$

Similarly the equation of the associate plane $\pi_n^{(v)}$ takes the form

$$(22) \quad 3F^2(y + nx) - n^2Dz = 0.$$

It can be shown that the associate plane $\pi_n^{(u)}[\pi_n^{(v)}]$ envelops a quadric cone which coincides with $\Gamma_2^{(u)}[\Gamma_2^{(v)}]$.

THEOREM. *The affine normals of the sections of $R^{(u)}[R^{(v)}]$ made by planes π_{np} at P lie in a plane $\pi_n^{(u)}[\pi_n^{(v)}]$. The envelope of this plane $\pi_n^{(u)}[\pi_n^{(v)}]$ is the associate cone $\Gamma_2^{(u)}[\Gamma_2^{(v)}]$.*

This may be taken as another construction of the $\Gamma_2^{(u)}[\Gamma_2^{(v)}]$. From this theorem we can easily write $\Gamma_2^{(u)}$ and $\Gamma_2^{(v)}$ in plane coordinates in the form

$$3F^2u_1u_3 + Au_2^2 = 0$$

and

$$3F^2u_2u_3 + D_1^2 = 0$$

respectively.

By eliminating ρ between (9) and (19) we obtain the locus of the osculating parabola (19) as a parabolic cylinder whose equation is of the form

$$(23) \quad z = (1/36n^3F^3)[3nF^2(y + nx) - Az]^2.$$

Similarly by replacing $R^{(u)}$ by $R^{(v)}$ we get another parabolic cylinder

$$(24) \quad z = (1/36n^3F^3)[3F^2(y + nx) - n^2Dz]^2.$$

It is easily seen that the diametral planes of (23) and (24) passing through P are (21) and (22) respectively. Hence we have the

THEOREM. *The locus of the osculating parabolas of the sections of $R^{(u)}[R^{(v)}]$ made by planes π_{np} is a parabolic cylinder whose diametral plane coincides with $\pi_n^{(u)}[\pi_n^{(v)}]$ and consequently envelops the cone $\Gamma_2^{(u)}[\Gamma_2^{(v)}]$.*

The above theorem is similar to a theorem of Kubota⁷ in which he has proved that the envelope of the Transon plane T , as t_n varies in a pencil, is a cone Γ_4 found by Su.

It is easily seen that the Transon plane T , the associate planes $\pi_n^{(u)}$ and $\pi_n^{(v)}$ belong to a pencil whose axis is the line t_n .

⁷ T. Kubota, "Einige Bemerkungen zur Affinflächen-theorie, *Science Reports Tohoku Imperial University* (1), vol. 19 (1930), pp. 163-168.

THEOREM. *The associate planes $\pi_n^{(u)}$ and $\pi_n^{(v)}$, and Transon plane T , are concurrent on the conjugate tangent of t_n .*

6. **Further associate cones.** The harmonic conjugate plane \bar{T} of the Transon plane T with respect to $\pi_n^{(u)}$ and $\pi_n^{(v)}$ has the equation

$$3F^2n(A + n^3D)(y + nx) - (A^2 + n^6D^2)z = 0,$$

or in plane coordinates

$$(25) \quad \begin{cases} \rho u_1 = 3F^2n^2(A + n^3D), \\ \rho u_2 = 3F^2n(A + n^3D), \\ \rho u_3 = -(A^2 + n^6D^2). \end{cases}$$

On eliminating ρ , n from (25) we obtain the equation of the envelope of \bar{T} in the form

$$A^2u_2 + D^2u_1^6 + 3F^2u_1u_2u_3(Au_2^3 + Du_1^3) = 0.$$

We state our result in the

THEOREM. *The harmonic conjugate plane \bar{T} of the Transon plane T with respect to the associate planes $\pi_n^{(u)}$ and $\pi_n^{(v)}$ envelops an algebraic cone of class six.*

We now consider the plane $\bar{\pi}$ conjugate to the plane π with respect to the associate planes $\pi_n^{(u)}$ and $\pi_n^{(v)}$. Hence the equation of $\bar{\pi}$ can be written in the form

$$3F^2(A + n^3D)(y + nx) - 2ADn^2z = 0$$

or

$$\begin{cases} \rho u_1 = 3F^2n(A + n^3D), \\ \rho u_2 = 3F^2(A + n^3D), \\ \rho u_3 = -2ADn^2. \end{cases}$$

The envelope of this plane as t_n varies is

$$3u_3(Au_2^3 + Du_1^3) + 2FJu_1^2u_2^2 = 0.$$

THEOREM. *The envelope of the plane $\bar{\pi}$, conjugate to the tangent plane π with respect to the associate planes $\pi_n^{(u)}$ and $\pi_n^{(v)}$ belonging to a variable tangent t_n , is an algebraic cone of class four.*

Now the equation of the harmonic conjugate plane $\bar{\omega}_1$ of π with respect to the associate plane $\pi_n^{(u)}$ and the Transon plane T can be written in the form

$$3nF^2(2A + n^3D)(y + nx) - 2A(A + n^3D)z = 0$$

or

$$\begin{cases} \rho u_1 = 3n^2F^2(2A + n^3D), \\ \rho u_2 = 3nF^2(2A + n^3D), \\ \rho u_3 = -2A(A + n^3D). \end{cases}$$

Hence the envelope of this plane has the equation

$$3F^2u_1u_3(2Au_2^3 + Du_1^3) + 2A_2^2(Au_2^3 + Du_1^3) = 0.$$

Similarly, the plane $\tilde{\omega}_2$ determined by

$$(\pi_n^{(v)}T\tilde{\omega}_2\pi) = -1$$

envelops a cone whose equation is

$$3F^2u_2u_3(Au_2^3 + 2Du_1^3) + 2Du_1^2(Au_2^3 + Du_1^3) = 0.$$

Hence we may state the

THEOREM. *The harmonic conjugate plane $\tilde{\omega}_1[\tilde{\omega}_2]$ of the tangent plane π with respect to the associated plane $\pi_n^{(u)}[\pi_n^{(v)}]$ and the Transon plane T envelops an algebraic cone of class five.*

7. Further consideration of the associate planes $\pi_n^{(u)}$ and $\pi_n^{(v)}$. From the last theorem of 5 we observe that the conjugate tangent t_n plays an especially important rôle in the affine theory of surfaces. The associate planes $\pi_{-n}^{(u)}$ and $\pi_{-n}^{(v)}$ belonging to the conjugate tangent t_n are worthy of consideration. The equations of $\pi_{-n}^{(u)}$ and $\pi_{-n}^{(v)}$ are

$$3nF^2(y - nx) + Az = 0$$

and

$$3F^2(y - nx) - n^2Dz = 0$$

respectively. Hence the line $l(\pi_n^{(u)}, \pi_{-n}^{(u)})$, the intersection of $\pi_n^{(u)}$ and $\pi_{-n}^{(u)}$, has the equations

$$\begin{cases} 3n^2F^2x - Az = 0, \\ y = 0 \end{cases}$$

and the line $l(\pi_n^{(v)}, \pi_{-n}^{(v)})$ has the equations

$$\begin{cases} 3n^2F^2y - Dz = 0, \\ x = 0. \end{cases}$$

Evidently as t_n varies $l(\pi_n^{(u)}, \pi_{-n}^{(u)})$ and $l(\pi_n^{(v)}, \pi_{-n}^{(v)})$ describe two pencils

with the center P . The plane in which $l(\pi_n^{(u)}, \pi_{-n}^{(u)})$ and $l(\pi_n^{(v)}, \pi_{-n}^{(v)})$ lie has the equation

$$(26) \quad 3n^2F^2(Dx + Ay) - ADz = 0.$$

This plane always passes through the tangent

$$z = Ay + Dx = 0.$$

As t_n varies the lines

$$l(\pi_n^{(u)}, \pi_{-n}^{(v)}) : \begin{cases} 6nF^2y - (A + n^3D)z = 0, \\ 6n^2F^2x - (A - n^3D)z = 0 \end{cases}$$

and

$$l(\pi_{-n}^{(u)}, \pi_n^{(v)}) : \begin{cases} 6nF^2x - (A + n^3D)z = 0, \\ 6n^2F^2y - (A - n^3D)z = 0 \end{cases}$$

describe the cones

$$162F^2x^2y^2 + 27(Ax^3 - Dy^3) - 9FJxyz^2 - J^2z^4 = 0$$

and

$$162F^2x^2y^2 + 27(Ay^3 - Dx^3) - 9FJxyz^2 - J^2z^4 = 0$$

respectively. These two cones are of order four and each has double contact with the tangent plane along the asymptotic tangents. The plane containing the lines $l(\pi_n^{(u)}, \pi_{-n}^{(v)})$ and $l(\pi_{-n}^{(u)}, \pi_n^{(v)})$ is

$$(27) \quad 6n^2F^2(x + y) - \{(A - n^3D) + n(A + n^3D)\}z = 0,$$

which describes a pencil with the tangent

$$z = x + y = 0$$

as its axis. It can be easily verified that the line determined by the planes (27) and (26) generates an algebraic cone Γ_7 of order seven.

8. Further considerations of diameters $d_n^{(u)}$ and $d_n^{(v)}$. The plane $\pi(d_n^{(u)}, d_n^{(v)})$, containing $d_n^{(u)}$ and $d_n^{(v)}$ is given by the equation

$$n(2A + n^3D)x + (A + 2n^3D)y - n^2FJz = 0$$

which envelops an algebraic cone of order six and class four. The equation of this cone in plane coordinates is

$$(28) \quad 4Ju_3(Au_2^3 + Du_1^3) + 18F^2Ju_1u_2u_3^2 + FJ^2u_1^2u_2^2 - 27F^3u_3^4 = 0.$$

The equation of the cone (28) in point coordinates is

$$27(A^2x^6 + D^2y^6) + 144F^3Jx^3y^3 + 54FJxyz(Ax^3 + Dy^3) \\ - J^2z^3(Ax^3 + Dy^3) - FJ^3xyz^4 + 15F^2J^2x^2y^2z^2 = 0.$$

Obviously the plane $\pi(d_{-n}^{(u)}, d_{-n}^{(v)})$, whose equation is

$$n(2A - n^3D)x - (A - 2n^3D)y + n^2FJz = 0,$$

also envelopes the cone (28). The line of intersection of $\pi(d_n^{(u)}, d_n^{(v)})$ and

$$\pi(d_{-n}^{(u)}, d_{-n}^{(v)}) \text{ is } \begin{cases} n^4Dx + Ay - n^2FJz = 0, \\ Dn^2y + Ax = 0 \end{cases}$$

which generates an algebraic cone Γ_3 of order three, whose equation is

$$Ax^3 + Dy^3 + FJxyz = 0.$$

This cone intersects the tangent plane in the tangents of Darboux.

In a similar way we may show that the planes $\pi(d_n^{(u)}, d_{-n}^{(v)})$ and $\pi(d_{-n}^{(u)}, d_n^{(v)})$ envelop the same cone whose equation is

$$4Ju_3(Au_2^3 + Du_1^3) - 6F^2Ju_1u_2u_3^2 - (5/3)FJu_1^2u_2^2 - 27F^3u_3^4 = 0,$$

and their line of intersection describes the cone

$$Ax^3 + Dy^3 + (5/3)FJxyz = 0.$$

Finally the planes $\pi(d_n^{(u)}, d_{-n}^{(u)})$ and $\pi(d_n^{(v)}, d_{-n}^{(v)})$ pass through the line

$$\begin{cases} Az + 3n^2F^2x = 0, \\ n^2Dz + 3F^2y = 0, \end{cases}$$

which describes the cone

$$(29) \quad 9F^4xy = ADz^2.$$

This cone has been studied first by Su⁸ and subsequently by Kubota.⁹ *The polar line of the tangent plane with respect to the cone (29) is the affine surface normal.*

NATIONAL UNIVERSITY OF CHEKIANG,
CHINA.

⁸ B. Su and A. Ichida, "On certain cones connected with a surface in affine space," *Japanese Journal of Mathematics*, vol. 10 (1930), p. 214.

⁹ T. Kubota, "Einige Bemerkungen zur Affinflächenentheorie," *Japanese Journal of Mathematics*, vol. 10 (1930), p. 216.

VORTICES AND NODES.*

By AUREL WINTNER.

Let a, b, c, d be real constants with a determinant

$$(1) \quad ad - bc \neq 0,$$

and let $f(x, y), g(x, y)$ be real-valued functions, defined in a circle, say $r \leq \alpha$, about the origin of an (x, y) -plane so as to be continuous and to satisfy

$$(2) \quad f(x, y) = o(r), \quad g(x, y) = o(r) \text{ as } r \rightarrow 0,$$

where

$$(3) \quad x = r \cos \theta, \quad y = r \sin \theta \quad (r > 0).$$

The following considerations, which will always assume that the above conditions are fulfilled, will deal with the problem initiated (and, in the analytic case, solved) by Poincaré (cf., e. g., [1]), namely, with the problem of asymptotic connections between the solutions of the system

$$(4) \quad x' = ax + by + f(x, y), \quad y' = cx + dy + g(x, y)$$

and of the trivial system

$$(5) \quad x' = ax + cy, \quad y' = cx + dy,$$

where the primes denote differentiations with respect to a real variable, t .

1. Let the point $(x, y) = (0, 0)$ be called an *attractor* of (4) if the above α can be replaced by a β ($\leq \alpha$) having the following property: If a solution path,

$$(6) \quad x = x(t), \quad y = y(t),$$

of (4) has at least one point in the circle $r < \beta$, then (6) tends to the point $(0, 0)$ (when *either* $t \rightarrow \infty$ *or* $t \rightarrow -\infty$). The existence of (6) for a whole t -half-line is part of this requirement. In Poincaré's terminology, vortices (foci) and nodes (of any kind) are attractors in this sense, but whirls (centra) and saddle points are not. Since f and g , instead of being analytic, are arbitrary continuous functions satisfying (2), an attractor can

* Received March 25, 1947.

be neither a pure vortex nor a pure node. In addition, (4) can have more than one solution (6) passing through the same point (x_0, y_0) , since the assumptions made before (3) do not imply anything like a Lipschitz condition. Nevertheless, the following fact is true under the assumptions preceding (3):

(i) *If $(0, 0)$ is an attractor of (5), then it is an attractor of (4) also.*

The converse of (i) is false, since, even in Poincaré's analytic case, $(0, 0)$ can be a vortex of (4) when it is a whirl of (5).

First, if a real matrix

$$(7) \quad \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

has a positive determinant, and if $B(x, y)$ denotes the bilinear form belonging to (7), it is easy to verify (for instance, by considering the various affine normal forms of a real binary matrix), that, after a suitable affine transformation of the (x, y) -space, either the characteristic numbers of (7) are purely imaginary or else

$$(8) \quad \min_{x^2+y^2=1} B(x, y) > 0$$

(whether the characteristic numbers be complex or real and, in the second case, whether (7) does or does not have a multiple elementary divisor). It follows that $(0, 0)$ is an attractor of (5) [if and] only if a matrix equivalent to the matrix of (5) satisfies (8). In fact, if the determinant of (7) is negative, then the characteristic numbers of (8) are real and of opposite sign, and so the affine normal form of (5) is $x' = px, y' = -qy$, where $p > 0$ and $q > 0$. Similarly, if (7) has purely imaginary characteristic numbers (and so, in view of (1), a positive determinant), the affine normal form of (5) is $x' = sy, y' = -sx$, where $s \neq 0$. Hence, $(0, 0)$ is a saddle point of (5) in the first case and a whirl of (5) in the second case, and so no attractor of (i) in either case. [It is clear, but for the present immaterial, that $(0, 0)$ is either a node or a vortex, and therefore an attractor, of (5) in the remaining cases.] Accordingly, (i) will be proved if it is shown that $(0, 0)$ must be an attractor of (4) if (7) satisfies (8).

To this end, let $r = r(t)$ in (6) refer to a solution (6) of (4). Since $xx' + yy' = rr'$, it is seen from (4) and from the definition of $B(x, y)$ that

$$rr' = B(x, y) + xf(x, y) + yg(x, y).$$

Hence, (2) and (8) imply that

$$rr' \geq \lambda r^2 + o(r^2),$$

where λ is a positive constant and $r = r(t)$. It will be convenient to replace t by $-t$ (this is admissible, since (4) does not contain t explicitly). Then it is clear from the last formula line that

$$r' \leq -\frac{1}{2}\lambda r \text{ whenever } r < \beta,$$

where λ and β are positive constants. But a well-known argument shows (cf. [3], Appendix) that the existence of two such positive constants λ, β implies the following fact: If a solution curve (6) of (5) is within the circle $r < \beta$ at *some* $t = t_0$, then this solution of (5) exists on the whole half-line $t_0 < t < \infty$ and is such that $r(t) \rightarrow 0$ holds as $t \rightarrow \infty$. Since this means that $(0, 0)$ is an attractor of (4), the proof of (i) is complete.

2. The term *node* was used above in its usual sense: The point $(0, 0)$ is a node of (4) if it is an attractor having the property that every solution path tending to $(0, 0)$ has a tangent at $(0, 0)$. Let $(0, 0)$ be called a *proper node* of (4) if it is a node and has, in addition, the property that every half-line issuing from $(0, 0)$ is the tangent at $(0, 0)$ of some solution path tending to $(0, 0)$.

In the trivial case (5), the point $(0, 0)$ is a node if (and only if) the characteristic numbers of (7) are real and of the same sign, but a proper node (if and) only if the characteristic numbers of (7) are equal and belong to distinct elementary divisors. [In fact, the prototype of (5) is $x' = x, y' = y$ in the latter case. This has the general solution $x = x_0 e^t, y = y_0 e^t$, which (besides the trivial solution determined by $x_0 = 0, y_0 = 0$) represents all half-lines issuing from $(0, 0)$. But there are two further cases of real characteristic numbers with common sign, namely, the cases in which the prototype of (5) becomes either $x' = x, y' = \lambda y$, where $0 < \lambda \neq 1$, or $x' = x, y' = x + y$. In these cases, the general solutions are $x = x_0 e^t, y = y_0 e^{\lambda t}$ and $x = x_0 e^t, y = (x_0 t + y_0) e^t$ respectively. Hence, $(0, 0)$ is a node in both cases, but is not a proper node in either case, since only a finite number of half-lines issuing from $(0, 0)$ become tangents in both cases.]

Since only the continuity of f and g is assumed in (2) and (5), nothing hinders that, when $(0, 0)$ is a proper node of (5), the correspondence between solution paths and half-lines through $(0, 0)$ be not one-to-one.

The following facts lie quite on the surface and are collected here only in order to contrast them with (iii) and (iv) below.

- (ii a) If $(0, 0)$ is a proper node of (5), it can be a vortex of (4).
- (ii b) If $(0, 0)$ is a vortex of (5), it must be a vortex of (4) also.

(ii c) If $(0, 0)$ is a vortex of (5), and if the unit of length on the t -axis is so chosen that the real part of the characteristic numbers of (7) becomes of absolute value 1, then $\log r(t) \sim -t$ holds for every solution (6) of (5) satisfying $r(t) = o(1)$.

Here, and in the sequel, the point $(0, 0)$ is called a *vortex* of (4) if it is an attractor of (4) and has the property that $|\theta(t)| \rightarrow \infty$ holds for every solution of (4) satisfying $0 < r(t) \rightarrow 0$; cf. (3) and (6).

Let $a = d = 1$ and $b = c = 0$. Then $(0, 0)$ is a proper node of (5), and (4) becomes

$$(9) \quad x' = x + f(x, y), \quad y' = y + g(x, y).$$

The assertion of (ii a) is proved by an example of Perron [2], pp. 128-129, as follows: Choose, in terms of (3),

$$f(x, y) = -h(r)r \sin \theta, \quad g(x, y) = h(r)r \cos \theta,$$

where $h = h(r)$ is any continuous function vanishing at $r = 0$. Then $f(x, y)$ and $g(x, y)$ are continuous functions satisfying (2). Clearly, $h(r)$ is identical with $(g(x, y) \cos \theta - f(x, y) \sin \theta)/r$ and therefore, by (3), with $(xg - yf)/r^2$ and so, by (9), with $(y'x - x'y)/r^2$, which is θ' , by (3). But the definition of f and g (in terms of an h) also shows that $xf + yg$ vanishes identically, and so $xx' + yy' = xx + yy$, by (9). In view of (3), this means that $r' = r$, that is, $r = r_0 e^t$. Since $h(r)$ was seen to be identical with θ' , it follows that $\theta(t) = \int h(r_0 e^t) dt$. Hence, it is sufficient to choose $h(r) = (\log r)^{-1}$ (if $r > 0$, and $h(0) = 0$), in order to see that $r(t) \rightarrow 0$ and $|\theta(t)| \rightarrow \infty$ (as $t \rightarrow -\infty$) are satisfied by every choice of both integration constants $\theta_0, r_0 (> 0)$. This proves (ii a).

The assertion of (ii b) is well-known and will be verified here only because a simple proof can be based on (i). In order to see this, let $(0, 0)$ be a vortex of (5). It can be assumed that (7) is in the normal form mentioned in (ii c). Then (5) appears in the form

$$(10) \quad x' = -x + \lambda y + f(x, y), \quad y' = -y - \lambda x + g(x, y)$$

(after a suitable rotation of the (x, y) -plane, the characteristic numbers of (7) being $-1 \pm i\lambda$, where $\lambda \neq 0$, and so, without loss of generality, $\lambda > 0$). Since θ' is identical with $(y'x - x'y)/r^2$, by (3), and therefore, by (10) and (2), with $1/r^2$ times $0 - \lambda x^2 - \lambda y^2 + o(r^2)$,

$$(11) \quad r^2 \theta' = -\lambda r^2 + o(r^2).$$

But $(0, 0)$ is a vortex of (5), hence an attractor of (5), and so, by (i), an attractor of (4). Hence, $r(t) \rightarrow 0$ holds when either $t \rightarrow \infty$ or $t \rightarrow -\infty$. As will be seen in the proof of (ii c) below, the normalization (10) implies that $r(t) \rightarrow 0$ holds when $t \rightarrow \infty$. Since $\lambda > 0$, it now follows from (11) that $\theta' = -\lambda + o(1)$, hence $\theta = -\lambda t + o(t)$, and so $|\theta| \rightarrow \infty$, as $t \rightarrow \infty$. This proves (ii b).

Finally, from (10) and (3),

$$(12) \quad rr' = -r^2 + xf + yg.$$

Hence, $r' = -r + o(r)$, by (2). Since this means that $(\log r)' = -1 + o(1)$ as $t \rightarrow \infty$, the assertion of (ii c) follows.

3. If (4) is of the form (10), then (5) becomes

$$(13) \quad x' = -x + \lambda y, \quad y' = -y - \lambda x$$

and has therefore the general solution

$$(14) \quad x = (v \cos \lambda t + u \sin \lambda t)e^{-t}, \quad y = (v \sin \lambda t - u \cos \lambda t)e^{-t},$$

where u and v are arbitrary integration constants. Since (14) implies that $r(t) = (x^2 + y^2)^{\frac{1}{2}}$ is identical with $r_0 e^{-t}$, where $r_0 = (u^2 + v^2)^{\frac{1}{2}}$ is arbitrary, one might expect that the (logarithmic) assertion of (ii c) can be refined to $r(t) \sim r_0 e^{-t}$, where $r_0 (> 0)$ is an integration constant. But the example proving (iii bis) below will show that this refinement of (ii c) is false. However, it becomes true under the Tauberian restriction which replaces the $o(r)$ in (2) by $O(r^{1+\epsilon})$, where $\epsilon > 0$. This Tauberian fact (and somewhat more) is contained in the following theorem:

(iii) If $(0, 0)$ is a vortex of (5) and if (2) in (4) is refined to

$$(15) \quad f(x, y) = O(r^{1+\epsilon}), \quad g(x, y) = O(r^{1+\epsilon}), \quad (\epsilon > 0),$$

then every solution path (6) of (4) tending to the vortex $(0, 0)$ of (4) is asymptotic to a solution path (6) of (5), and every solution path (6) of (5) is asymptotic to a solution path (6) of (4).

It is not claimed that this asymptotic correspondence between the solutions of (4) and those of the trivial system (5) is a one-to-one correspondence. In fact, such a claim is prevented, among other things, by the circumstance that the assumptions placed by (iii) on f and g , assumptions consisting of (15) and of the mere continuity of f and g , are compatible with a variety of

solution paths of (5) which pass through the same point $(x_0, y_0) \neq (0, 0)$; cf. the remarks made before (i).

What will actually be ascertained is an integral refinement of the order restriction (15):

(iii*) *The assertions of (iii) remain true if (15) is relaxed to*

$$(16) \quad |f(x, y)| \leq \psi(r) \quad |g(x, y)| \leq \psi(r),$$

where $\psi(r)$ is any continuous, monotone non-decreasing function of r defined for $0 \leq r \leq \alpha$ in such a way that

$$(17) \quad \psi(r) = o(r) \quad (r \rightarrow 0)$$

and

$$(18) \quad \int_0^\alpha r^{-2} \psi(r) dr < \infty.$$

On the other hand, some restriction of the o in (2) is indispensable:

(iii bis) *The assertions of (iii) can fail if (15) is relaxed to (2).*

In order to see this, let, in virtue of (3),

$$f(x, y) = h(r)r \cos \theta, \quad g(x, y) = h(r)r \sin \theta,$$

where $h(r)$ is a continuous function vanishing at $r = 0$. Then f and g are continuous functions satisfying (2). But (12) now becomes $r' = -r + rh(r)$. In particular, if $h(r) = 1/\log r$, then $w' = -1 + 1/w$, where $w = \log r$. This differential equation for w gives $w + \log(w - 1) = t_0 - t$, that is, $r(\log r - 1) = r_0 e^{-t}$. Hence, the assertion of (iii), which implies that $r(t) \sim r_0 e^{-t}$, where $r_0 (> 0)$ is arbitrary, cannot be true in this case. This proves (iii bis).

4. In the proof of (iii) and of its generalization (iii*), it can be assumed that (4) and (5) are in their respective normal forms, (10) and (13).

Let u and v in (14) be thought of as functions of t , to be determined in such a way that (14) becomes a solution (6) of (10) (variation of constants). To this end, the formal substitution of (14) into (10) supplies the necessary and sufficient conditions

$$(19) \quad u' = (f \cos \lambda t + g \sin \lambda t)e^t, \quad v' = (f \sin \lambda t - g \cos \lambda t)e^t,$$

where $f = f(x, y)$ and $g = g(x, y)$ must be expressed, in terms of (14), as

functions of u, v and t . In other words, (19) represents two differential equations of the form

$$(20) \quad u' = f^*(u, v, t), \quad v' = g^*(u, v, t),$$

where f^* and g^* are continuous functions defined in the region

$$(21) \quad -\infty < t < \infty, \quad (u^2 + v^2)^{\frac{1}{2}} e^{-t} \leq \alpha$$

of the (u, v, t) -space.

By the asymptotic correspondences referred to in (iii) and (iii*) is meant the following: If (6) is any solution of (10) satisfying $r(t) = o(1)$, as $t \rightarrow \infty$, then the corresponding solution

$$(22) \quad u = u(t), \quad v = v(t)$$

of (20) is such that the limits

$$(23) \quad \lim_{t \rightarrow \infty} u(t) = u(\infty), \quad \lim_{t \rightarrow \infty} v(t) = v(\infty)$$

exist; conversely, if a pair of constants $u(\infty), v(\infty)$ is arbitrarily specified, then there exists at least one solution (22) of (20) satisfying (23).

According to (19), the functions f^*, g^* occurring in (20) are majorized by $(f^2 + g^2)^{\frac{1}{2}} e^t$. Hence, if the factor $2^{\frac{1}{2}}$ is thought of as being submerged into the function sign ψ , then, by (14) and (16),

$$|f^*(u, v, t)| \leq \psi((u^2 + v^2)^{\frac{1}{2}} e^{-t}) e^t, \quad |g^*(u, v, t)| \leq \psi((u^2 + v^2)^{\frac{1}{2}} e^{-t}) e^t,$$

if (u, v, t) is a point in the region (21). Let C denote an arbitrary positive number and let $T = T(C)$ be defined by $Ce^{-T} = \alpha$. Then, by the monotony of ψ ,

$$|f^*(u, v, t)| \leq \psi(Ce^{-t}) e^t \quad |g^*(u, v, t)| \leq \psi(Ce^{-t}) e^t,$$

if (u, v, t) is a point in the region

$$(24) \quad u^2 + v^2 \leq C^2, \quad T \leq t < \infty.$$

Thus, it is possible to define a pair of functions $F(u, v, t), G(u, v, t)$ which are continuous in the product space of the entire (u, v) -plane and the half-line $0 \leq t < \infty$, are given by

$$(25) \quad F(u, v, t) = f^*(u, v, t), \quad G(u, v, t) = g^*(u, v, t)$$

when (u, v, t) is a point of the region (24), and satisfy the inequality

$$(26) \quad (F(u, v, t)^2 + G(u, v, t)^2)^{\frac{1}{2}} \leq \lambda(t)$$

for all points (u, v, t) , where

$$(27) \quad \lambda(t) = \begin{cases} \psi(Ce^{-t})e^t & \text{if } Ce^{-t} \leq \alpha, \\ \psi(\alpha)e^t & \text{if } Ce^{-t} > \alpha. \end{cases}$$

In order to apply to the system of differential equations

$$(28) \quad u' = F(u, v, t), \quad v' = G(u, v, t)$$

the theorem proved in [3], let the majorant, $\lambda(t)$, on the right of (26) be written in the form $\lambda(t)\phi(u^2 + v^2)$, where $\phi(r) \equiv 1$. Then

$$(29) \quad \int_0^\infty \lambda(t) dt < \infty \quad \text{and} \quad \int_1^\infty dr/\phi(r) = \infty,$$

since

$$\int_0^\infty \lambda(t) dt = \int_0^T \psi(\alpha)e^t dt + \int_T^\infty \psi(Ce^{-t})e^t dt$$

and

$$\int_T^\infty \psi(Ce^{-t})e^t dt = C \int_0^a \psi(r)r^{-2} dr < \infty,$$

by (18).

It now follows from the general theorem of [3] that, if (22) is any solution of (28), then the limits (23) exist, and that the latter can attain arbitrary values $u(\infty)$, $v(\infty)$ when (22) is suitably chosen.

On the other hand, since (25) holds on the region (24), it is clear that, if (22) is any solution of (30) which satisfies

$$(30) \quad u(t)^2 + v(t)^2 \leq C^2 \text{ for sufficiently large } t,$$

then it is a solution of (20) for all large t , and conversely. Since the constant C was arbitrary, it follows that the proof of (iii*) can be completed by showing that, if (6) is any solution of (10) satisfying $r(t) = o(1)$, then, for the corresponding solution (22) of (20), there exists a constant C satisfying (30). In view of (14), this is equivalent to the assertion

$$(31) \quad r(t)e^t = O(1), \quad (t \rightarrow \infty).$$

If $r > 0$ is small enough, then, according to (12) and (16),

$$r' \leq -r + 2\psi(r).$$

On the other hand, (17) shows that the inequality $r - 2\psi(r) > 0$ holds whenever $r > 0$ is small enough. Since $r = r(t)$ tends to 0 as $t \rightarrow \infty$, it follows that, when t is large enough,

$$(r - 2\psi(r))^{-1}r' < -1.$$

Finally, the identity

$$(r - 2\psi(r))^{-1} = r^{-1} + 2\psi(r)(r^2 - 2r\psi(r))^{-1},$$

when integrated with respect to t , supplies the inequality

$$r \leq \text{const. } e^{-t} \exp \left(2 \int_r^\infty \psi(u)(u^2 - 2u\psi(u))^{-1} du \right), \quad r = r(t).$$

Clearly, (31) now follows from (17) and (18).

This completes the proof of (iii) and of its generalization (iii*).

5. The preceding proof has nowhere used the assumption that $\lambda \neq 0$ in (10), (13), (14). But if $\lambda = 0$, then (13) and (14) become $x' = -x$, $y' = -y$ and $x = x_0 e^{-t}$, $y = y_0 e^{-t}$ respectively, representing (except for the unit of length and the orientation on the t -axis) the only case in which $(0, 0)$ is a proper node of (5); cf. the second section in 2 above. Correspondingly, it is clear from the definition of a proper node that the assertions of (iii) now become equivalent to the statement that $(0, 0)$ is a proper node of (4). Accordingly, the proof of (iii) implies the following theorem:

(iv) *If $(0, 0)$ is a proper node of (5) and if (2) in (4) is refined to (15), then $(0, 0)$ is a proper node of (4) also.*

This limiting case of (iii) was proved by Perron ([2], p. 123, Fall II) under the assumption of continuous partial derivatives f_x, f_y, g_x, g_y for the functions $f(x, y), g(x, y)$ in a circle about $(0, 0)$. In (iv), these derivatives need not exist at any $(x, y) \neq (0, 0)$, the functions f, g being just continuous. Correspondingly, Perron's additional statement that just one solution path of the system (4) issues from $(0, 0)$ in every direction is not true under the general assumption of (iv), since the situation is the same as in the observation following (iii).

Perron's method of proof in his particular case of (iv) consists in a reduction of the system (4) to a single equation $dy/dx = F(x, y)$. Such a reduction is made possible, of course, only by the circumstance that $x = x(t)$ or $y = y(t)$ in (4) becomes substantially monotone near the node $(0, 0)$. Since this circumstance is not presented by the spirals about a vortex $(0, 0)$, Perron's method would not lead to (iii) even under the assumption of continuous partial derivatives for f, g (cf. Perron [2], pp. 280-283, where only (ii b) is proved for vortices).

(iv*) *The assertion of (iv) remains true if (15) is relaxed to (16), (17), (18).*

In fact, (iv*) relates to (iii*) in the same way as (iv) to (iii).

However, *some* Tauberian restriction of the o in (2) is necessary in order that the assertion of (iv) be correct. In fact, the truth of the negation which corresponds to (iii bis) is a corollary of (ii a).

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES

- [1] J. Horn, *Gewöhnliche Differentialgleichungen beliebiger Ordnung*, Sammlung Schubert, vol. 50 (1905), §75-§76.
- [2] O. Perron, Ueber die Gestalt der Integralkurven einer Differentialgleichung erster Ordnung in der Umgebung eines singulären Punktes, *Mathematische Zeitschrift*, vol. 15 (1922), pp. 121-146 and vol. 16 (1923), pp. 273-295.
- [3] A. Wintner, "Asymptotic integration constants," *American Journal of Mathematics*, vol. 68 (1946), pp. 553-559.

ANALYTICITY IN HILBERT SPACE AND SELF-ADJOINT TRANSFORMATIONS.*

By MARTINUS ESSER.

1. Introduction. The spectral theory of self-adjoint transformations in Hilbert space has been developed in many different ways, both for particular cases and for the general case [von Neumann],¹ [Stone] etc. One method, which has formerly been used only for two particular types of self-adjoint transformations [Hellinger I and II], is based on Cauchy's integral theorem for analytic functions, and obtains the spectrum of the transformation by studying the singularities of certain analytic functions. The present paper will show how this method can be used to derive the whole theory of self-adjoint transformations. To this end, we must synthesize the usual features of analyticity and of Hilbert space, and use the notion of analytic dependency of elements in Hilbert space on a complex variable. We begin by studying integrals of elements in Hilbert space. Such integrals will be used for contour integrations of analytic elements and in the formulation of our final result [cf. Maeda and, later, Riesz and Lorch].

2. Integrals in Hilbert space. We consider a Hilbert space H . To distinguish between points $f(\lambda)$ in H and complex numbers $L(\lambda)$, both depending on a complex variable λ , we shall ascribe the word "element" to the first case and restrict the word "function" to the second case. The inner product of two elements f, g will be denoted by (f, g) and the modulus $(f, f)^{1/2}$ will be denoted by $|f|$. Definitions of equality, limits, continuity, series and integrals in the space H can be made in the usual manner by means of the modulus $|f|$. Two types of Stieltjes integrals will be considered, namely

$$(1) \quad \int_{\Gamma} f(\lambda) dL(\lambda) = \lim_{\delta \rightarrow 0} \sum_n f(\lambda'_n) \Delta_n L(\lambda)$$

and

$$(2) \quad \int_{\Gamma} L(\lambda) df(\lambda) = \lim_{\delta \rightarrow 0} \sum_n L(\lambda'_n) \Delta_n f(\lambda),$$

* Received November 9, 1946. The present article is derived from the first part of my doctoral dissertation [Esser]. The dissertation has been written under the guidance of Ernst Hellinger, Northwestern University.

¹ Names in brackets refer to references.

where $f(\lambda)$ is an element in H , $L(\lambda)$ a complex valued function and λ a complex variable describing a curve Γ . These integrals have properties similar to those of ordinary Stieltjes integrals, and we shall assume such properties without proof. In particular the integral (1) exists when the element $f(\lambda)$ is continuous and the variation of $L(\lambda)$ on Γ is finite. A similar sufficient condition of existence can be formulated for the integral (2) if we define the variation of an element as follows:

Definition I. The variation of an element $f(\lambda)$ on a curve Γ is defined to be the least upper bound of $|\sum_n a_n \Delta_n f(\lambda)|$ for all possible partitions of the curve Γ and for any complex numbers a_n whose modulus does not exceed one.

It should be noted that the analogies between the variation of functions and the variation of elements are not as complete as the analogies between integrals of functions and integrals of elements. For instance the variation of an element $f(\lambda)$ does not equal, in general, the least upper bound of $\sum |\Delta_n f(\lambda)|$, and the variation of an element over the sum of two intervals of λ may differ from the sum of the variations over each interval.

We have the following theorem:

THEOREM II. If an element $f(x)$ depending on a real variable x satisfies the orthogonality relation

$$(3) \quad (f(b) - f(a), f(d) - f(c)) = 0$$

for any set of numbers $a \leq b \leq c \leq d$, then the variation of $f(x)$ on any interval (y, z) is finite and equals $|f(z) - f(y)|$.

Proof. We divide the interval (y, z) into successive intervals and denote by $\Delta_n f$ the increment of $f(x)$ over the n -th interval. We consider also complex numbers a_n with $|a_n| \leq 1$. We have then, by the linear properties of inner products,

$$(4) \quad \left(\sum_n a_n \Delta_n f, \sum_n a_n \Delta_n f \right) = \sum_{n,m} a_n \bar{a}_m (\Delta_n f, \Delta_m f).$$

By the orthogonality relation (3), and by the definition of moduli of elements, relation (4) becomes $|\sum_n a_n \Delta_n f|^2 = \sum_n |a_n|^2 |\Delta_n f|^2$. Considering different sets of a_n , we obtain the least upper bound of the second member by taking each a_n equal to one, and looking at the first member, we see that this upper bound is $|\sum \Delta_n f|^2 = |f(z) - f(y)|^2$. This value is independent of the sub-intervals of (y, z) considered, and therefore equals the square of the variation of $f(x)$ on that interval.

3. Analyticity in Hilbert space. A limit in H of the form

$$\frac{df(\lambda)}{d\lambda} = \lim_{\mu \rightarrow \lambda} \frac{f(\mu) - f(\lambda)}{\mu - \lambda},$$

where $f(\lambda)$ is an element in H and λ is a complex variable, will be called a derivative in H . An element $f(\lambda)$ which is defined and has a derivative for each λ of an open domain D in the λ plane, will be said to be analytic over D .

The usual properties of analytic functions can be generalized to analytic elements [Cf. Wiener]. We have the following theorems:

Let $f(\lambda)$ be an analytic element over a domain D . Then:

THEOREM III. *For any element g , the function $(f(\lambda), g)$ is analytic over D and has the derivative $(df(\lambda)/d\lambda, g)$.*

THEOREM IV. *For any two points a, b in D , the integral $\int_a^b f(\lambda) d\lambda$ exists and is independent of the curve of integration joining a to b , provided that such curves are not separated by points not belonging to D .*

THEOREM V. *Considering a curve Γ in D encircling once and in the positive sense a point λ , we have Cauchy's formula*

$$2\pi i f(\lambda) = \int_{\Gamma} (\mu - \lambda)^{-1} f(\mu) d\mu.$$

THEOREM VI. *For each point μ in D , we have the Taylor series expansion*

$$f(\lambda) = \sum_{n=0}^{\infty} (1/n!) (\lambda - \mu)^n (d^n f(\mu)/d\mu^n),$$

convergent over any circle centered at μ and contained in D .

Only Theorems III and IV will be used in this article. The integral in Theorem IV exists for each curve joining a to b because $f(\lambda)$ is continuous, and the integral is independent of the path because, for each element g , the number $(\int_a^b f(\lambda) d\lambda, g) = \int_a^b (f(\lambda), g) d\lambda$ is independent of the path.

4. Self-adjoint transformations. The analytic element with which we shall be concerned in this article is the element whose existence is stated by the following theorem.

THEOREM VII. *Considering a self-adjoint transformation T , an arbitrary element u in H , and a non-real complex variable λ , the relation*

$$(5) \quad T[f(\lambda)] - \lambda f(\lambda) = u$$

defines a unique element $f(\lambda)$. This element is analytic and its derivative satisfies the relation

$$(6) \quad T[df(\lambda)/d\lambda] - \lambda df(\lambda)/d\lambda = f(\lambda).$$

Denoting by λ_2 the imaginary part of λ , we have the inequality

$$(7) \quad |f(\lambda)| \leq |\lambda_2|^{-1} |u|.$$

The existence and uniqueness of $f(\lambda)$, and the inequality (7) are well-known [Stone, Definition 2.11, Theorem 4.14]. The existence of the derivative can be shown as follows: The transformation T being linear, we have

$$(8) \quad T[f(\lambda) - f(\mu)] - \lambda[f(\lambda) - f(\mu)] = (\lambda - \mu)f(\mu).$$

Inequality (7) gives then

$$|f(\lambda) - f(\mu)| \leq |\lambda_2|^{-1} |\lambda - \mu| |f(\mu)| \leq |\lambda_2 \mu_2|^{-1} |\lambda - \mu| |u|.$$

The right member of this inequality approaches zero when μ approaches λ , therefore $f(\lambda)$ is continuous.

We divide both members of (8) by $(\lambda - \mu)$:

$$(9) \quad T \left[\frac{f(\lambda) - f(\mu)}{\lambda - \mu} \right] - \lambda \frac{f(\lambda) - f(\mu)}{\lambda - \mu} = f(\mu),$$

and let μ approach λ . The second member approaches $f(\lambda)$. In the first member $[f(\lambda) - f(\mu)]/[\lambda - \mu]$ must then approach a limit, because the transformation from $f(\mu)$ to $[f(\lambda) - f(\mu)]/[\lambda - \mu]$ defined by equation (9) is a closed bounded transformation [Stone, Definitions 2.5 and 2.13]. Therefore $df(\lambda)/d\lambda$ exists and satisfies relation (6). Theorem VII is thus proved.

We shall now find a few properties of the element $f(\lambda)$ which will justify the consideration of Lemma VIII below.

We consider elements satisfying the two relations

$$(10) \quad \begin{aligned} T[f(\lambda)] - \lambda f(\lambda) &= u, \\ T[g(\mu)] - \mu g(\mu) &= v. \end{aligned}$$

By definition of self-adjoint transformations, we have $(T[f(\lambda)], g(\mu)) = (f(\lambda), T[g(\mu)])$. Substituting for $T[f]$ and $T[g]$ the values given by (10), and using the linear properties of inner products, we get

$$(11) \quad (\lambda - \bar{\mu})(f(\lambda), g(\mu)) = (f(\lambda), v) - (u, g(\mu)),$$

where $\bar{\mu}$ denotes the conjugate of μ .

We introduce a function $L(\lambda)$ defined by

$$(12) \quad L(\lambda) = (f(\lambda), u) = \overline{(u, f(\lambda))}.$$

By Schwarz's inequality and inequality (7), we have

$$(13) \quad |L(\lambda)| \leq |f(\lambda)| |u| \leq |\lambda_2|^{-1} |u|^2.$$

By taking $v = u$, $g(\mu) = f(\mu)$ in equation (11) and using equation (12), we obtain

$$(14) \quad (\lambda - \bar{\mu})(f(\lambda), f(\mu)) = L(\lambda) - \overline{L(\mu)}.$$

5. Proof of a lemma. Relations (13) and (14) lead to a lemma whose proof constitutes the essential part of our construction of the spectrum of the transformation T .

LEMMA VIII. *Let $f(\lambda)$ be an analytic element defined for non-real λ , and satisfying for each non-real λ and μ the relation*

$$(15) \quad (\lambda - \bar{\mu})(f(\lambda), f(\mu)) = L(\lambda) - \overline{L(\mu)},$$

where $L(\lambda)$ is a function for which $|\lambda_2 L(\lambda)|$ stays bounded. Then there exists an element $u(x)$ depending on the real variable x , which has a finite variation on the interval $-\infty \leq x \leq +\infty$, and such that

$$(16) \quad f(\lambda) = \int_{-\infty}^{+\infty} du(x)/(x - \lambda).$$

The element $u(x)$ is continuous on the right

$$(17) \quad u(x) = \lim_{\epsilon \rightarrow 0} u(x + \epsilon), \quad \epsilon > 0,$$

and satisfies the orthogonality relation

$$(18) \quad (u(b) - u(a), u(d) - u(c)) = 0$$

for any set of numbers $a \leq b \leq c \leq d$.

The function $L(\lambda)$ has the three following properties:

1). $L(\lambda)$ is analytic in the upper half λ plane, as is seen by Theorem III applied to equation (15).

2). $L(\lambda)$ has a non-negative imaginary part $L_2(\lambda)$ for positive λ_2 , because, for $\lambda = \mu$, relation (15) becomes $\lambda_2 |f(\lambda)|^2 = L_2(\lambda)$.

3). By hypothesis, $\lambda_2 |L(\lambda)|$ is bounded.

It is known [Doob and Koopman, Esser] that functions $L(\lambda)$ which have the three preceding properties can be represented, for $\lambda_2 > 0$, by the integral

$$(19) \quad L(\lambda) = \int_{-\infty}^{+\infty} (x - \lambda)^{-1} d\rho(x),$$

where $\rho(x)$ is a certain real, bounded, non-decreasing function. We may suppose $\rho(x)$ continuous on the right.

Moreover, when we make $\lambda = \bar{\mu}$ in equation (15), we see that $\overline{L(\lambda)} = L(\bar{\lambda})$, and therefore relation (19) is also valid when λ is in the lower half plane. Substituting the integral (19) for L in equation (15), we obtain successively

$$(20) \quad (\lambda - \bar{\mu})(f(\lambda), f(\mu)) = \int_{-\infty}^{+\infty} \frac{d\rho(x)}{x - \lambda} - \int_{-\infty}^{+\infty} \frac{d\rho(x)}{x - \bar{\mu}},$$

$$(21) \quad (f(\lambda), f(\mu)) = \int_{-\infty}^{+\infty} \frac{d\rho(x)}{(x - \lambda)(x - \bar{\mu})}.$$

This formula shows us how the inner product $(f(\lambda), f(\mu))$ depends on the variables λ, μ , and thereby will enable us to study convergence properties of the element $f(\lambda)$.

We consider two points a, b in the same half λ plane, and integrate the two members of (21) with respect to λ over a curve from a to b in that half λ plane. In the first member, the integral of the inner product may be replaced by the inner product of the integral of $f(\lambda)$ by $f(\mu)$. In the second member, the order of the two integrations can be interchanged because of the uniform convergence of the Stieltjes integral. We thus obtain

$$\left(\int_a^b f(\lambda) d\lambda, f(\mu) \right) = \int_{-\infty}^{+\infty} \int_a^b \frac{d\lambda}{x - \lambda} \frac{d\rho(x)}{x - \bar{\mu}}.$$

We rewrite this equation with a and b replaced by \bar{b} and \bar{a} , and add the two equations. Defining, except for an additive element independent of λ , an element $u(\lambda)$ by

$$(22) \quad 2\pi i[u(b) - u(a)] = \int_a^b f(\lambda) d\lambda + \int_{\bar{b}}^{\bar{a}} f(\lambda) d\lambda,$$

we obtain

$$(23) \quad \pi(u(b) - u(a), f(\mu)) = \int_{-\infty}^{+\infty} \frac{\arg \frac{a - x}{b - x}}{x - \bar{\mu}} \frac{d\rho(x)}{x - \bar{\mu}},$$

where \arg stands for the argument contained between $-\pi$ and $+\pi$.

By repeating the sequence of operations which transformed (21) into (23), we transform (23) into

$$(24) \quad \pi^2(u(b) - u(a), u(d) - u(c)) = \int_{-\infty}^{+\infty} \arg \frac{a-x}{b-x} \arg \frac{c-x}{d-x} d\rho(x),$$

where a, b, c, d are any four numbers with positive imaginary parts. If, in particular, we take $c = a, d = b$, we obtain

$$(25) \quad \pi^2 |u(b) - u(a)|^2 = \int_{-\infty}^{+\infty} \arg^2 \frac{a-x}{b-x} d\rho(x).$$

The function $\rho(x)$, being monotone, is continuous for almost all x . When a and b approach a real point y where $\rho(x)$ is continuous, then the integral in (25) approaches zero. Therefore $u(a)$ converges when a approaches y , and we can define an element $u(y)$ by $u(y) = \lim_{a \rightarrow y} u(a)$. If we let a approach a real point y of continuity of $\rho(x)$, the members of equations (23), (24) and (25) approach limits, which are obtained by replacing a by y in these members. Similarly, we can let b, c, d approach real points where $\rho(x)$ is continuous, and thus find that the equations (23), (24), (25) stay valid when a, b, c, d are real points where $\rho(x)$ is continuous. The formulas so obtained are:

$$(26) \quad (u(b) - u(a), f(\mu)) = \int_a^b \frac{d\rho(x)}{x - \mu}, \quad a \leq b,$$

$$(27) \quad (u(b) - u(a), u(d) - u(c)) = 0, \quad a \leq b \leq c \leq d,$$

$$(28) \quad |u(b) - u(a)|^2 = \rho(b) - \rho(a), \quad a \leq b.$$

When a and b approach from the right an arbitrary real point x , the members of (28) approach zero. Therefore $\lim_{\epsilon \rightarrow 0} u(x + \epsilon)$, where $x + \epsilon$ varies over the points of continuity of $\rho(x)$ on the right of x , exists for all x . This limit will define $u(x)$ at the points of discontinuity of $\rho(x)$.

Thus we have defined for all real x an element $u(x)$, which will be the element $u(x)$ mentioned in Lemma VIII. This element satisfies the relation (17). Moreover, by letting one or more of the points a, b, c, d approach, from the right, points of discontinuity of $\rho(x)$, we see that formulas (26), (27) and (28) remain valid when one or more of the points a, b, c, d are points of discontinuity of $\rho(x)$. Thus in particular the orthogonality relation (18) is proved.

Since $\rho(x)$ is bounded and monotone, the limits $\rho(-\infty) = \lim_{x \rightarrow -\infty} \rho(x)$ and $\rho(+\infty) = \lim_{x \rightarrow +\infty} \rho(x)$ exist. Therefore, by (28), the limits $u(-\infty) = \lim_{x \rightarrow -\infty} u(x)$ and $u(+\infty) = \lim_{x \rightarrow +\infty} u(x)$ exist. The element $u(x)$ has been determined hitherto except for an additional element independent of x . This element may be determined by the additional condition

$$(29) \quad u(-\infty) = \lim_{x \rightarrow -\infty} u(x) = 0.$$

By Theorem II, the variation of $u(x)$ on the infinite interval of x is finite. It equals $[\rho(+\infty) - \rho(-\infty)]^{\frac{1}{2}}$.

The integral of equation (16) exists because $(x - \lambda)^{-1}$ is continuous and the variation of $u(x)$ is finite. By the construction used in its definition, the integral is contained in the closed linear manifold determined by the values of the element $f(\mu)$ considered for all non-real μ . Therefore, to prove equation (16), it is sufficient to show that its two members have equal projections on each element $f(\mu)$, in other words that for all non-real μ we have

$$(30) \quad (f(\lambda), f(\mu)) = \left(\int_{-\infty}^{+\infty} \frac{du(x)}{x - \lambda}, f(\mu) \right).$$

Using (26) and (21), we get successively

$$\begin{aligned} \left(\int_{-\infty}^{+\infty} \frac{du(x)}{x - \lambda}, f(\mu) \right) &= \int_{-\infty}^{+\infty} \frac{1}{x - \lambda} d(u(x), f(\mu)) \\ &= \int_{-\infty}^{+\infty} \frac{1}{x - \lambda} d \int_{-\infty}^x \frac{d\rho(y)}{y - \bar{\mu}} = \int_{-\infty}^{+\infty} \frac{d\rho(x)}{(x - \lambda)(x - \bar{\mu})} = (f(\lambda), f(\mu)), \end{aligned}$$

which proves (30) and therefore (16). The proof of Lemma VIII is thus completed.

6. Spectral projections. Given a self-adjoint transformation T , the correspondence of elements $f(\lambda)$ to elements u established by relation (5), and of elements $u(x)$ to elements $f(\lambda)$ established in Section 5, associates with each element u in H an element $u(x)$ defined and unique for each real x . This element $u(x)$ will be called the "spectral projection" of u . In the present section, we shall show mainly that for each real x , $u(x)$ is the projection of u on a linear manifold \mathfrak{M}_x , this manifold depending only on x and T , but not on u .

We shall first show that

$$(31) \quad u(+\infty) = \lim_{x \rightarrow +\infty} u(x) = u.$$

Equations (26), (29), (19) and (12) give successively

$$(u(+\infty), f(\mu)) = \int_{-\infty}^{+\infty} \frac{d\rho(x)}{x - \bar{\mu}} = \overline{L(\mu)} = (u, f(\mu)).$$

Therefore, and because $u(+\infty)$ is in the closed linear manifold determined by $f(\mu)$ for all non-real μ , we find that $u(+\infty)$ is the projection of u on

this manifold. To establish (31), we have yet to show that u is in this manifold.

If u is in the domain \mathcal{D} of T , we get successively

$$T\{T[f(\lambda)]\} - \lambda T[f(\lambda)] = T[u] \quad \text{by (5),}$$

$$|T[f(\lambda)]| \leq |\lambda_2|^{-1} |T[u]| \quad \text{by (7),}$$

$$\lim_{\lambda_2 \rightarrow +\infty} T[f(\lambda)] = 0,$$

$$u = \lim_{\lambda_2 \rightarrow +\infty} \lambda f(\lambda) \quad \text{by (5).}$$

The last equation shows that u is in the closed linear manifold determined by $f(\lambda)$ for all non-real λ , and therefore $u = u(+\infty)$. This equality can be extended to arbitrary elements u in H because the domain \mathcal{D} of T is everywhere dense in H and the transformation from u to $u(+\infty)$ is linear and bounded. Therefore (31) is proved.

We show next that the orthogonality relation (18) can be generalized into

$$(32) \quad (u(b) - u(a), v(d) - v(c)) = 0, \quad a \leq b \leq c \leq d,$$

where $u(x)$, $v(x)$ are the spectral projections of two different elements u , v . To these elements u , v correspond elements $f(\lambda)$, $g(\mu)$ defined by equations (10). By (16) we get

$$(33) \quad (f(\lambda), v) = \int_{-\infty}^{+\infty} \frac{d\sigma(x)}{x - \lambda},$$

where $\sigma(x) = (u(x), v)$. The function $\sigma(x)$ has a finite variation on the infinite interval because the variation of the element $u(x)$ is finite, and it is continuous on the right because $u(x)$ is continuous on the right. By taking $\lambda = \bar{\mu}$ in equation (11), we see that

$$(u, g(\mu)) = (f(\bar{\mu}), v) = \int_{-\infty}^{+\infty} \frac{d\sigma(x)}{x - \bar{\mu}}.$$

Therefore equation (11) becomes

$$(\lambda - \bar{\mu})(f(\lambda), g(\mu)) = \int_{-\infty}^{+\infty} \frac{d\sigma(x)}{x - \lambda} - \int_{-\infty}^{+\infty} \frac{d\sigma(x)}{x - \bar{\mu}}.$$

From this equation we can derive (32) by the method used in deriving (18) from (20).

By particularizing a , b , c , d in (32), and using (29) and (31), we get $(u(x), v - v(x)) = 0$ for each x . Thus, for each x , the linear manifold \mathfrak{M}_x formed by all elements $u(x)$ is orthogonal to the linear manifold formed

by all elements $u - u(x)$. As each element u in H is the sum of an element $u(x)$ in \mathfrak{M}_x and an element orthogonal to \mathfrak{M}_x , the manifold \mathfrak{M}_x must be closed and $u(x)$ is the projection of u on \mathfrak{M}_x .

The orthogonality relation (32) gives, moreover, $(u(x), v - v(y)) = 0$ for $x \leq y$. Therefore $u(x)$ belongs to the manifold \mathfrak{M}_y of all elements orthogonal to all $v - v(y)$, and we infer that \mathfrak{M}_x is contained in \mathfrak{M}_y .

7. The transformation T . This section will prove that an element f , with spectral projection $f(x)$, is in the domain \mathcal{D} of the transformation T if, and only if, the integral $\int_{-\infty}^{+\infty} x df(x)$ exists, and that we have then

$$(34) \quad T[f] = \int_{-\infty}^{+\infty} x df(x).$$

Let us first suppose that f is an element in \mathcal{D} . We define an element u by

$$(35) \quad u = T[f] - if.$$

Then formula (16) gives $f = \int_{-\infty}^{+\infty} (y - i)^{-1} du(y)$. Projecting on \mathfrak{M}_x , we get $f(x) = \int_{-\infty}^x (y - i)^{-1} du(y)$. From this integral we obtain, for each real a and b , $u(b) - u(a) = \int_a^b (x - i) df(x)$. The first member has the limit u when $a \rightarrow -\infty$, $b \rightarrow +\infty$. Therefore $\int_{-\infty}^{+\infty} (x - i) df(x)$ exists and we have

$$(36) \quad u = \int_{-\infty}^{+\infty} (x - i) df(x) = \int_{-\infty}^{+\infty} x df(x) - if.$$

Comparing (35) and (36), we obtain (34).

Conversely, let f be such that the integral in (34) exists. Then we define u by (36). We obtain $u(x) = \int_{-\infty}^x (y - i) df(y)$. The integral $\int_{-\infty}^{+\infty} (x - i)^{-1} du(x)$ exists and equals f . Therefore f is the element which satisfies (35). It results that f is in \mathcal{D} and satisfies again (34).

8. Conclusion. We have shown how the theory of analytic functions can be used to derive the known characterization of self-adjoint transformations, which is:

THEOREM IX. *Each self-adjoint transformation T in Hilbert space can be defined as follows by means of a certain family of closed linear manifolds*

\mathfrak{M}_x : Let $f(x)$ denote the projection of f on \mathfrak{M}_x . Then f is in the domain of T if and only if the integral $\int_{-\infty}^{+\infty} x df(x)$ exists. The integral equals then $T[f]$.

The manifolds \mathfrak{M}_x are such that:

- 1). For $x \leq y$, \mathfrak{M}_x is a subset of \mathfrak{M}_y .
- 2). For each f we have $\lim_{x \rightarrow -\infty} f(x) = 0$ and $\lim_{x \rightarrow +\infty} f(x) = f$.
- 3). The element $f(x)$ is continuous on the right: $f(x) = \lim_{\epsilon \rightarrow 0} f(x + \epsilon)$, $\epsilon > 0$.

ILLINOIS INSTITUTE OF TECHNOLOGY.

REFERENCES.

-
- J. Doob and B. Koopman, "On analytic functions with positive imaginary parts," *Bulletin of the American Mathematical Society*, vol. 40 (1934), pp. 601-605.
- M. H. M. Esser, "Spectral expansion for self-adjoint differential transformations and for arbitrary self-adjointly transformations," Dissertation, Northwestern University (February, 1946).
- E. Hellinger, I, "Neue Begründung der Theorie quadratischer Formen von unendlichvielen Veränderlichen," *Crelle's Journal für Mathematik*, vol. 136 (1909), pp. 210-271.
- E. Hellinger, II, "Zur Stieltjesschen Kettenbruchtheorie," *Mathematische Annalen*, vol. 86 (1922), pp. 18-29.
- F. Maeda, Several articles from 1934 on, reviewed in *Zentralblatt für Mathematik*, vols. 8, 11, 12, 13 and 14.
- J. von Neumann, "Allgemeine Eigenwerttheorie Hermitescher Funktionaloperatoren," *Mathematische Annalen*, vol. 102 (1929), pp. 49-131.
- F. Riesz and E. Lorch, "The integral representation of unbounded self-adjoint transformations in Hilbert space," *Transactions of the American Mathematical Society*, vol. 39 (1936), pp. 331-340.
- M. Stone, "Linear transformations in Hilbert space," *American Mathematical Society, Colloquium publications*, vol. 15, 1932.
- N. Wiener, "Note on a paper of M. Banach," *Fundamenta Mathematicae*, vol. 4 (1923), pp. 136-143.

ON THE SUMMABILITY OF MULTIPLE FOURIER SERIES.*

By A. ZYGMUND.

1. Let $f(x, y)$ be an L -integrable function of period 2π with respect to both x and y . Let

$$(1) \quad \sum_{m,n=-\infty}^{+\infty} c_{mn} e^{i(mx+ny)}$$

be the Fourier series of f , so that

$$c_{mn} = 1/(2\pi)^2 \int_0^{2\pi} \int_0^{2\pi} f(x, y) e^{-i(mx+ny)} dx dy$$

for all m and n . As regards the convergence of the first arithmetic means

$$(2) \quad \begin{aligned} \sigma_{m,n}(x, y) &= \sigma_{m,n}(x, y; f) \\ &= \sum_{\mu, \nu=-m, -n}^{m, n} (1 - (|\mu|/(m+1)))(1 - (|\nu|/(n+1))) e^{i(\mu x + \nu y)} \end{aligned}$$

of the series (1) the following results are known.

THEOREM A. *There is an f such that $\sigma_{m,n}(x, y)$ diverges everywhere. More precisely,*

$$\limsup_{m,n \rightarrow +\infty} \sigma_{m,n}(x, y) = +\infty$$

everywhere.

THEOREM B. *If m and n tend to $+\infty$ in such a way that the ratios m/n and n/m remain bounded, then*

$$\sigma_{m,n}(x, y; f) \rightarrow f(x, y)$$

almost everywhere.

THEOREM C. *If not only f but also $|f| \log^+ |f|$ is integrable, then $\sigma_{m,n}(x, y; f) \rightarrow f(x, y)$ almost everywhere as m and n tend to $+\infty$ independently of each other.*

Corresponding results hold for the Abel means

$$f_{r,\rho}(x, y) = \sum_{m,n=-\infty}^{+\infty} c_{mn} e^{i(mx+ny)} r^{|m|} \rho^{|n|}$$

* Received February 1, 1947.

of (1) as $r, \rho \rightarrow 1$. The condition of boundedness of the ratios m/n and n/m is then to be replaced by the boundedness of the ratios $(1-r)/(1-\rho)$ and $(1-\rho)/(1-r)$.

All this, and the literature, will be found in Saks [5], Zygmund [7], Jessen, Marcinkiewicz and Zygmund [2], and Marcinkiewicz and Zygmund [3]. It may be added that considering the limit of σ_{mn} when m and n tend to infinity in such a way that m/n and n/m are bounded, was first suggested by C. N. Moore [4, p. 567]. That the last restriction is essential, follows from Theorem A. Of course, this kind of summability—let us call it *restricted summability* (C, 1)—is analogous to the classical concept of *restricted differentiability* of multiple integrals. For, as shown by Lebesgue, for any integrable $f(x, y)$,

$$\lim_{h, k \rightarrow 0} 1/hk \int_0^h \int_0^k f(x+u, y+v) du dv = f(x, y)$$

almost everywhere, provided h and k tend to 0 in such a way that the ratios h/k and k/h are bounded.

The main purpose of this note is to prove the following extension of Theorem B.

THEOREM 1. *Let $m = m(t)$ and $n = n(t)$ be any two non-decreasing and integer valued functions of the parameter t , $0 \leq t < \infty$, tending to infinity with t . Then, for any integrable $f(x, y)$*

$$\lim_{t \rightarrow +\infty} \sigma_{m,n}(x, y; f) = f(x, y)$$

at almost every point (x, y) . More generally, let λ be any number ≥ 1 . Then

$$\sigma_{\mu, \nu}(x, y; f) \rightarrow f(x, y)$$

almost everywhere, if μ and ν tend to infinity in such a way that

$$(3) \quad \lambda^{-1}m(t) \leq \mu \leq \lambda m(t), \quad \lambda^{-1}n(t) \leq \nu \leq \lambda n(t).$$

The second part of this theorem reduces to the first part if $\lambda = 1$. If $m(t) = n(t) = [t]$ ($[t]$ = the integer part of t), we obtain Theorem B. For clearly, if Theorem 1 is proved for any fixed number $\lambda \geq 1$, it is automatically established for $\lambda = \lambda(x, y)$ varying from point to point. The novelty of Theorem 1, in comparison with Theorem B, is that m and n need no longer be of the same order of magnitude (consider, for example σ_{n, n^2} or $\sigma_{n, 2^n}$). Their rate of increase may be totally different, provided that—

except for a bounded factor—one of them is a monotone function of the other.

The analogue of (the second part of) Theorem 1 for Abel means is as follows.

THEOREM 2. *Let $\phi(u)$ and $\psi(u)$ be non-decreasing functions of u defined in the interval $0 < u \leq 1$, satisfying the inequalities $0 < \phi(u) \leq 1$, $0 < \psi(u) \leq 1$, and tending to 0 with u . Let $\lambda \geq 1$ be fixed. Then, for every integrable f and almost every point (x, y) ,*

$$f_{r,\rho}(x, y) \rightarrow f(x, y)$$

as r and ρ tend to 1 in such a way that

$$(4) \quad \begin{aligned} \lambda^{-1}\phi(u) &\leq 1 - r \leq \lambda\phi(u), \\ \lambda^{-1}\psi(u) &\leq 1 - \rho \leq \lambda\psi(u). \end{aligned}$$

More generally, at almost every point (x, y) we have

$$f_{r,\rho}(\xi, \eta) \rightarrow f(x, y)$$

provided the points re^{it} and $\rho e^{i\eta}$ tend respectively to e^{ix} and e^{iy} along any non-tangential paths and in such a way that conditions (4) are satisfied.

It will be sufficient here to prove Theorem 1 only. The proof of Theorem 2 is completely analogous since the Poisson and the Féjér kernels satisfy similar inequalities (see inequalities (4) below), which are their only properties required in the proof. The fact that r and ρ are, unlike m and n , continuous variables does not affect the argument. Nor does the case of non-tangential paths mentioned in Theorem 2 introduce any new difficulty. For all that, Theorem 2 is more interesting in applications than Theorem 1. We shall return to some of these applications in another note.

The proof of Theorem 1 is given in Section 2, below. Section 3 will be devoted to some additional results.

2. LEMMA 1. *Let $h(t)$ and $k(t)$ be two positive functions defined for $t > 0$, non-decreasing, and tending to 0 with t . Let E be any plane set whose outer measure $|E|$ is finite and positive. Suppose that to every point $(x, y) \in E$ corresponds a rectangle $R = R_{x,y}$ with center (x, y) , and sides $2h(t)$, $2k(t)$ parallel to the axes, where $t = t(R)$ varies with R . Then there is a finite number of rectangles $R_{x_0y_0}, R_{x_1y_1}, \dots, R_{x_ny_n}$ without points in common and such that*

$$(5) \quad \sum_{m=0}^n |R_m| > |E|/26$$

where $R_m = R_{x_m y_m}$.

Proof. Let K_0 denote the aggregate of all the rectangles R corresponding to the points of E . Let

$$(6) \quad t^*_0 = \sup_{R \in K_0} t(R).$$

We may assume that the sides of the rectangle $R \in K_0$ are bounded. For otherwise there would exist rectangles R with areas arbitrarily large, and (5) could obviously be satisfied with $n = 0$.

Let us now define a rectangle R_0 and a number $t_0 = t(R_0)$ by the following conditions. If t^*_0 in (6) is actually attained, that is if there is a rectangle $R \in K_0$ such that $t^*_0 = t(R)$, we take that R for R_0 , and set $t_0 = t^*_0$. Otherwise, we take for R_0 any R such that $t_0 = t(R)$ satisfies both conditions

$$h(t_0) \geq \frac{1}{2}h(t^*_0 - 0), \quad k(t_0) \geq \frac{1}{2}k(t^*_0 - 0).$$

Let us now denote by K'_1 the set of all the rectangles $R \in K_0$ which have points in common with R_0 , and let K_1 be the class of the remaining rectangles R . Thus $K_0 = K'_1 + K_1$. Moreover, it is immediately seen that if we denote by \tilde{R}_0 the rectangle concentric with R_0 , with sides parallel to the axes and dimensions five times those of R_0 , then all the rectangles $R \in K'_1$ are covered by \tilde{R}_0 .

Let us now set

$$(7) \quad t^*_1 = \sup_{R \in K_1} t(R),$$

and let us define a rectangle R_1 and a number $t_1 = t(R_1)$ by the following conditions. If t_1 in (7) is attained, we take $t_1 = t^*_1$, and for R_1 we take the corresponding R . Otherwise, we take for R_1 any $R \in K_1$ such that $t_1 = t(R)$ satisfies the conditions

$$h(t_1) \geq \frac{1}{2}h(t^*_1 - 0), \quad k(t_1) \geq \frac{1}{2}k(t^*_1 - 0).$$

Thus R_1 has no points in common with R_0 . Let K'_2 be the set of all the rectangles $R \in K_1$ which have points in common with R_1 , and K_2 the set of the remaining rectangles from K_1 . Hence $K_1 = K'_2 + K_2$. Again, all rectangles from K'_2 are covered by the rectangle \tilde{R}_1 concentric with and similar to R_1 , with sides five times larger.

The general procedure is now clear. Suppose we have already defined

$t_{m-1}^*, t_{m-1}, K_{m-1}, R_{m-1}$. We then set $K_{m-1} = K'_m + K_m$, where K'_m consists of the rectangles $R \in K_{m-1}$ which have points in common with R_{m-1} , and K_m of the remaining rectangles from K_{m-1} . The rectangles from K'_m are contained in the rectangle \bar{R}_{m-1} concentric with, similar to, and five times larger than R_{m-1} . We set

$$(8) \quad t_m^* = \sup_{t \in K_m} t(R)$$

and set $t_m = t_m^*$ if the supremum in (8) is attained. Otherwise we take $t_m = t(R)$ for an $R \in K_m$ and satisfying

$$h(t_m) \geq \frac{1}{2}h(t_m^* - 0), \quad k(t_m) \geq \frac{1}{2}k(t_m^* - 0).$$

Obviously R_m has no point in common with R_{m-1}, \dots, R_1, R_0 .

The sequence R_0, R_1, R_2, \dots may be finite or not. In the former case, K_m is empty for some m . Let us first suppose that the sequence is infinite. Since $t_0^* \geq t_1 \geq t_2 \geq \dots$, there are two possibilities

- (i) all the numbers t_m^* are bounded below by a positive number;
- (ii) the numbers t_m^* tend to 0.

In case (i), inequality (5) is obvious for n large enough. Let us therefore pass to case (ii). It is easy to see that every rectangle R' from K_0 is contained in some \bar{R}_m . For suppose that this is not true. That would mean that for each m the rectangle R' is contained in K_m , which is clearly impossible since the dimensions of the rectangles from K_m do not exceed $2h(t_m^*)$, $2k(t_m^*)$, and so tend to 0 with $1/m$.

Since E is contained in the sum of the rectangles R from K_0 , it must be contained in $\bar{R}_0 + \bar{R}_1 + \dots$. Hence,

$$(9) \quad |E| \leq \sum_{m=0}^{\infty} |\bar{R}_m| = 25 \sum_{m=0}^{\infty} |R_m|,$$

and this gives (5) for n large enough.

If the sequence $\{R_m\}$ is finite and ends with R_n , the above argument gives (9) with ∞ replaced by n . The inequality (5) is then true *a fortiori*.

Remark. The above proof is a simple adaptation of the proof in the special case $h(t) = t, k(t) = \alpha t$, α being any positive but fixed number (see Marcinkiewicz and Zygmund [3]). It is easy to see that the coefficient $1/26$ in (5) could be replaced by any number $> 1/9$, but the numerical value of

it is without importance. There is an extension of Lemma 1 to the case when each point of E belongs to infinitely many rectangles R with dimensions tending to 0 (see Jessen, Marcinkiewicz and Zygmund [2]). This extension is not needed here.

LEMMA 2. Let $f(x, y)$ be an integrable function defined in the square

$$(Q') \quad -2\pi \leq x \leq 2\pi, \quad -2\pi \leq y \leq 2\pi,$$

and let $h(t)$ and $k(t)$ be the functions of Lemma 1. For (x, y) belonging to the square

$$(Q) \quad -\pi \leq x \leq \pi, \quad -\pi \leq y \leq \pi$$

let

$$(10) \quad f_*(x, y) = \sup_t \frac{1}{4hk} \int_{-h}^h \int_{-k}^k |f(x+u, y+v)| \, du \, dv,$$

where t is so small that the rectangle over which we integrate is contained in Q' . For any $\xi > 0$, let $\mathcal{E}_*(\xi)$ denote the set of points $(x, y) \in Q$ at which $f_*(x, y) > \xi$. Then

$$(11) \quad |\mathcal{E}_*(\xi)| \leq 26\xi^{-1} \iint_{Q'} |f(x, y)| \, dx \, dy.$$

Proof. If $(x, y) \in \mathcal{E}_*(\xi)$, there is a rectangle R with center (x, y) , with sides parallel to the axes and of length $2h(t)$, $2k(t)$. By Lemma 1, we can select a finite number of these rectangles without points in common and such that the set $E = \mathcal{E}_*(\xi)$ satisfies (5). This gives

$$\iint_{Q'} |f| \, dx \, dy \geq \sum_{m=0}^n \iint_{R_m} |f| \, dx \, dy > \xi |\mathcal{E}_*(\xi)| / 26,$$

from which (11) follows.

LEMMA 3. Let $h(t)$, $k(t)$, $f(x, y)$ be the functions of Lemma 2, let α and β be fixed positive numbers, and let $f^{*\alpha, \beta}(x, y)$ be the functions $f_*(x, y)$ of Lemma 2 with $h(t)$, $k(t)$ replaced by $\alpha h(t)$ and $\beta k(t)$ respectively. For $(x, y) \in Q$, let

$$f^{**}(x, y) = \sup_{i,j} \{f_*^{2^i, 2^j}(x, y) 2^{-\frac{1}{2}(i+j)}\} \quad \text{for } i, j = 0, 1, 2, \dots$$

and let $\mathcal{E}^*(\xi)$ be the set of points $(x, y) \in Q$ at which $f^{**}(x, y) > \xi$. Then

$$(12) \quad |\mathcal{E}^*(\xi)| \leq A\xi^{-1} \iint_{Q'} |f| \, dx \, dy.$$

Proof. Let $\mathcal{E}_*^{\alpha, \beta}(\xi)$ be the set $\mathcal{E}_*(\xi)$ of Lemma 2 when we replace there $h(t)$, $k(t)$ by $\alpha h(t)$, $\beta k(t)$. A necessary and sufficient condition for the inequality $f^*(x, y) > \xi$ is that $f_*^{2^i, 2^j}(x, y) > \xi 2^{(i+j)/2}$ for some non-negative integers i, j . Thus

$$\mathcal{E}^*(\xi) \subset \sum_{i, j=0}^{\infty} \mathcal{E}_*^{2^i, 2^j}(\xi 2^{\frac{1}{2}(i+j)}),$$

and so

$$|\mathcal{E}^*(\xi)| \leq \sum_{i, j=0}^{\infty} |\mathcal{E}_*^{2^i, 2^j}(\xi 2^{\frac{1}{2}(i+j)})| \leq 26\xi^{-1} \left(\sum_{i, j=0}^{\infty} 2^{-\frac{1}{2}(i+j)} \right) \iint_Q |f| dx dy,$$

which leads to (12).

The arithmetic means $\sigma_{\mu\nu}$ of the Fourier series of f are given by the formula

$$(13) \quad \sigma_{\mu\nu}(x, y) = \pi^{-2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} f(x+u, y+v) K_{\mu}(u) K_{\nu}(v) du dv$$

where $K_{\mu}(u)$ is the Fejér kernel,

$$K_{\mu}(u) = \frac{1}{\mu+1} \frac{\sin^2 \frac{1}{2}(\mu+1)u}{2 \sin^2 \frac{1}{2}u}$$

satisfying the inequalities

$$(14) \quad K_{\mu}(u) \leq A_{\mu}, \quad K_{\mu}(u) \leq A\mu^{-1}u^{-2}$$

for $\mu \geq 1$ and $0 \leq u \leq \pi$. By A we mean here and hereafter a positive absolute constant not necessarily always the same.

Let

$$(15) \quad \bar{\sigma}_{\mu\nu}(x, y) = \pi^{-2} \int_0^{\pi} \int_0^{\pi} |f(x+u, y+v)| K_{\mu}(u) K_{\nu}(v) du dv$$

and let $m = m(t)$ and $n = n(t)$ be the functions of Theorem 1. Then

$$\begin{aligned} \pi^2 \bar{\sigma}_{\mu\nu}(x, y) &= \int_0^{1/m} du \int_{1/m}^{\pi} \{ \} dv + \int_0^{1/n} dv \int_{1/m}^{\pi} \{ \} du \\ &\quad + \int_{1/m}^{\pi} du \int_{1/n}^{\pi} \{ \} dv + \int_0^{1/m} du \int_0^{1/n} \{ \} dv \end{aligned}$$

where the curly brackets $\{ \}$ stand for the integrand on the right of (15). Using the inequalities (14) and observing that for each t

$$\lambda^{-1} \leq \mu/m \leq \lambda, \quad \lambda^{-1} \leq \nu/n \leq \lambda,$$

we immediately find that

$$\begin{aligned} 0 \leq \bar{\sigma}_{\mu\nu}(x, y) &\leq A\lambda^2 mn^{-1} \int_0^{1/m} du \int_{1/n}^{\pi} v^{-2} |f(x+u, y+v)| dv \\ &\quad + A\lambda^2 m^{-1} n \int_0^{1/n} dv \int_{1/m}^{\pi} u^{-2} |f(x+u, y+v)| du \\ &\quad + A\lambda^2 m^{-1} n^{-1} \int_{1/m}^{\pi} \int_{1/n}^{\pi} u^{-2} v^{-2} |f(x+u, y+v)| du dv \\ &\quad + A\lambda^2 mn \int_0^{1/m} \int_0^{1/n} |f(x+u, y+v)| du dv \\ &= A\lambda^2 P_t(x, y) + A\lambda^2 Q_t(x, y) + A\lambda^2 R_t(x, y) + A\lambda^2 S_t(x, y), \end{aligned}$$

say.

Let

$$*P(x, y) = \sup_t P_t(x, y) = \sup_t mn^{-1} \int_0^{1/m} du \int_{1/n}^{\pi} v^{-2} |f(x+u, y+v)| dv$$

and let us similarly define $*Q(x, y)$, $*R(x, y)$, $*S(x, y)$.

LEMMA 4. For $(x, y) \in Q$, each of the functions $*P$, $*Q$, $*R$, $*S$ is majorized by $Af^*(x, y)$, where f^* is the function of Lemma 3 formed with $h(t) = 1/m(t)$, $k(t) = 1/n(t)$.

Proof. Let us consider the integers $I = I(t)$, $J = J(t)$ defined by the conditions

$$\pi \leq 2^I/m < 2\pi, \quad \pi \leq 2^J/n < 2\pi.$$

Thus

$$\begin{aligned} P_t(x, y) &= mn^{-1} \sum_{j=1}^J \int_0^{1/m} dv \int_{2^{j-1}/n}^{2^j/n} v^{-2} |f(x+u, y+v)| dv \\ &\leq 4mn \sum_{j=1}^J 2^{-2j} \int_0^{1/m} du \int_{2^{j-1}/n}^{2^j/n} |f(x+u, y+v)| dv \leq 16 \sum_{j=1}^J 2^{-j} f^{1, 2^j}(x, y), \end{aligned}$$

where the function $f^{a, \beta}(x, y)$ is formed with $h(t) = 1/m(t)$, $k(t) = 1/n(t)$. From the definition of $f^*(x, y)$ it follows immediately that

$$P_t(x, y) \leq 16f^*(x, y) \sum_1^{\infty} 2^{-j/2}, \quad *P(x, y) \leq Af^*(x, y).$$

It is clear that the last inequality holds for $*Q(x, y)$. As to $*R$, we observe that

$$\begin{aligned}
R_t(x, y) &\leq m^{-1}n^{-1} \sum_{i,j=1}^{I,J} \int_{2^{i-1}/m}^{2^i/m} \int_{2^{j-1}/n}^{2^j/n} u^{-2}v^{-2} |f(x+u, y+v)| \, du dv \\
&\leq 16mn \sum_{i,j=1}^{I,J} 2^{-2(i+j)} \int_{2^{i-1}/m}^{2^i/m} \int_{2^{j-1}/n}^{2^j/n} |f(x+u, y+v)| \, du dv \\
&\leq 64 \sum_{i,j=1}^{I,J} 2^{-(i+j)} f_{*,2^i,2^j}(x, y) \\
&\leq 64 f^*(x, y) \sum_{i,j=1}^{\infty} 2^{-(i+j)/2},
\end{aligned}$$

so that $*R(x, y) \leq Af^*(x, y)$. Since $*S(x, y) \leq 4f_*(x, y) \leq 4f^*(x, y)$, Lemma 4 is proved.

Let us introduce the function

$$(16) \quad \sigma_{\lambda}^*(x, y; f) = \sup_{\mu, \nu} |\sigma_{\mu\nu}(x, y; f)|,$$

where $\mu \geq 1$ and $\nu \geq 1$ satisfy (3). From the formulas (13) and (15) we see that $|\sigma_{\mu\nu}|$ is majorized by a sum of four integrals typical of which is $\bar{\sigma}_{\mu\nu}$. Since

$$\bar{\sigma}_{\mu\nu}(x, y; f) \leq A\lambda^2 \{ *P(x, y) + *Q(x, y) + *R(x, y) + *S(x, y) \},$$

Lemma 4 and Lemma 3 lead to the following

LEMMA 5. *The function $\sigma_{\lambda}^*(x, y)$ is majorized by $A\lambda^2 f^*(x, y)$, where f^* is the same as in Lemma 4. The set of points $(x, y) \in Q$ at which $\sigma_{\lambda}^*(x, y; f) > \xi > 0$ is of measure not exceeding*

$$A\lambda^2 \xi^{-1} \iint_Q |f| \, dx dy.$$

Theorem 1 is a corollary of Lemma 5. For let us consider a decomposition

$$(17) \quad f = f_1 + f_2,$$

where $f_1(x, y)$ is a trigonometric polynomial, and the integral $\iint_Q |f_2| \, dx dy$ is arbitrarily small. Given an arbitrary $\delta > 0$, we may assume, by virtue of Lemma 2, that the set of points (x, y) at which $\sigma_{\lambda}^*(x, y; f_2)$ exceeds δ is of measure less than δ . Since $\sigma_{\mu\nu}(x, y; f_1)$ tends (uniformly) to $f_1(x, y)$, and since $|\sigma_{\mu\nu}(x, y; f_2)|$ is less than δ outside a set of measure $< \delta$, it follows without difficulty that $\sigma_{\mu\nu}(x, y; f) = \sigma_{\mu\nu}(x, y; f_1) + \sigma_{\mu\nu}(x, y; f_2)$ tends to $f(x, y)$ almost everywhere, provided conditions (3) are satisfied.

3. For certain applications one needs extensions of Theorems 1 and 2

to Fourier-Stieltjes series, that is to series (1) whose coefficients are represented by Stieltjes integrals

$$c_{mn} = (2\pi)^{-2} \int \int_Q e^{-i(mx+ny)} dF(E).$$

Here, as before, Q denotes the square $-\pi \leq x \leq \pi$, $-\pi \leq y \leq \pi$, and $F(E)$ is an additive function of sets. The arithmetic means of the series (1) are then represented by the formula

$$(19) \quad \sigma_{\mu\nu}(x, y; dF) = \pi^{-2} \int \int_Q K_\mu(x-u) K_\nu(y-v) dF(E).$$

As is very well known, the function $F(E)$ has almost everywhere a finite derivative $f(x, y)$, and $f(x, y)$ is also the derivative of the absolutely continuous component of $F(E)$.

THEOREM 3. *Let $F(E)$ be an additive function of sets in Q , and let $f(x, y)$ be the derivative of F . Then, at almost every point*

$$\sigma_{\mu\nu}(x, y; dF) \rightarrow f(x, y)$$

under the same conditions for μ and ν as in Theorem 1.

In the same sense, Theorem 2 remains valid for Fourier-Stieltjes series.

It is again sufficient to prove the part of Theorem 3 concerning the arithmetic means. A perusal of the proofs of Lemmas 2, 3, 4, 5 shows that these remain valid in the new case. In particular (compare Lemma 5) the set of points $(x, y) \in Q$ at which $\sigma_{\lambda}^*(x, y; dF) > \xi > 0$ is of measure not exceeding

$$(20) \quad A\lambda^2\xi^{-1} \int \int_Q |dF(E)|.$$

This immediately shows that at almost every point the numbers $\sigma_{\mu\nu}(x, y; dF)$ are bounded. To prove, however, that they tend to $f(x, y)$ we have to adopt a method slightly different from the one used before, since no decomposition corresponding to (17) can be used for singular mass distributions.

Let $F = F_1 + F_2$ be the decomposition of F into its absolutely continuous and singular parts. Thus $\sigma_{\mu\nu}(x, y; dF) = \sigma_{\mu\nu}(x, y; dF_1) + \sigma_{\mu\nu}(x, y; dF_2)$. Since $\sigma_{\mu\nu}(x, y; dF_1) = \sigma_{\mu\nu}(x, y; f) \rightarrow f(x, y)$ almost everywhere, it is enough to prove that $\sigma_{\mu\nu}(x, y; dF_2) \rightarrow 0$ almost everywhere. We first prove the following lemma.

LEMMA 6. *Suppose that the function F of Theorem 3 has the property*

that $\int_R |dF(E)| = 0$, where R is a rectangle $\alpha < x < \alpha'$, $\beta < y < \beta'$. Then at almost every point of R we have $\sigma_{\mu\nu}(x, y; dF) \rightarrow 0$ as μ, ν tend to $+\infty$ independently of each other.

Proof. In the formula (19) we can now integrate over the set $Q - R$. If $(x, y) \in R$, there is an $\eta > 0$ such that for every $(u, v) \in Q - R$ at least one of the inequalities $|x - u| \geq \eta$, $|y - v| \geq \eta$ is satisfied. Since $K_\mu(t)$ tend uniformly to 0 if $\eta \leq |t| \leq \pi$, it follows that

$$\begin{aligned} & |\sigma_{\mu\nu}(x, y; dF)| \\ & \leq o(1) \int \int_{Q-R} K_\mu(x-u) |dF(E)| + o(1) \int \int_{Q-R} K_\nu(y-v) |dF(E)| \\ & \leq o(1) \left[\int \int_Q K_\mu(x-u) |dF(E)| + \int \int_Q K_\nu(y-v) |dF(E)| \right], \end{aligned}$$

and it is enough to show that the first of the integrals in square brackets is bounded for almost every x , and the second for almost every y . It suffices to consider the first integral. It can be written

$$\int_{-\pi}^{\pi} K_\mu(x-u) d\chi(u),$$

if $\chi(u)$ denotes the integral $\int \int |dF|$ extended over the rectangle $-\pi \leq x \leq u$, $-\pi \leq y \leq \pi$. The last integral is the $(C, 1)$ mean of the Fourier-Stieltjes series of $d\chi(u)$, and so is bounded (indeed, tends to a limit) for almost every x . This completes the proof of Lemma 6.

Let us now revert to the singular function $F_2(E)$. It is well known that, given any number $\epsilon > 0$, we can find an open set $O \subset Q$, of measure differing from that of Q as little as we please and such that $\int \int_O |dF_2(E)| < \epsilon$. Let us write $F_2(E) = F_2(OE) + F_2(E - O) = F_3(E) + F_4(E)$, say. By Lemma 6, $\sigma_{\mu\nu}(x, y; dF_4)$ converges to 0 almost everywhere in O . The set of points of Q , and *a fortiori* the set of points of O , at which $\sigma_{\mu\nu}^*(x, y; dF_3) > \epsilon^{1/2}$, is of measure

$$\leq A\lambda^2\epsilon^{-1/2} \int \int_Q |dF_3| = A\lambda^2\epsilon^{-1/2} \int \int_O |dF_2| < A\lambda^2\epsilon^{1/2}.$$

Thus, if we exclude from O a subset of measure $< A\lambda^2\epsilon^{1/2}$, at the remaining points of O the least upper bound of the numbers $|\sigma_{\mu\nu}(x, y; dF_2)|$ is $< \epsilon$. Since both ϵ and $|Q - O|$ can be arbitrarily small, $\sigma_{\mu\nu}(x, y; dF_2)$ tends to 0 almost everywhere in Q . This completes the proof of Theorem 3.

From the extension of Lemma 5 to Fourier-Stieltjes series (cf. 20)), we see that the function $\sigma^*_\lambda(x, y; dF)$ is integrable in any positive power less than 1. More precisely,

THEOREM 4. *Under the assumptions of Theorem 3, and for $0 < p < 1$,*

$$(21) \quad \left\{ \int_Q \int_Q [\sigma^*_\lambda(x, y; dF)]^p dx \right\}^{1/p} \leq (A/(1-p))\lambda^2 \int_Q |dF|$$

$$(22) \quad \int_Q \int_Q [|\sigma_{\mu\nu}(x, y; dF) - f(x, y)|^p dx] \rightarrow 0.$$

The corresponding result holds for Abel means.

So far we have discussed, for simplicity, the case of double Fourier series. The results corresponding to Theorems 1, 2, 3, 4 hold, however, for Fourier series (or Fourier-Stieltjes series)

$$(23) \quad \sum_{n_1, \dots, n_k = -\infty}^{+\infty} c_{n_1, \dots, n_k} e^{i(n_1 x_1 + \dots + n_k x_k)}$$

of functions of k variables. So does, as is well known (see Jessen, Marcinkiewicz and Zygmund [2]), Theorem B, which then asserts that if not only the function $f(x_1, \dots, x_k)$ is integrable, but also $|f|(\log^+ |f|)^{k-1}$, then the (C. 1) means

$$\begin{aligned} & \sigma_{n_1, \dots, n_k}(x_1, x_2, \dots, x_k; f) \\ &= \sum_{\nu_1, \dots, \nu_k = -n_1, \dots, -n_k}^{n_1, \dots, n_k} \left(1 - \frac{|\nu_1|}{n_1 + 1}\right) \cdots \left(1 - \frac{|\nu_k|}{n_k + 1}\right) e^{i(\nu_1 x_1 + \dots + \nu_k x_k)} \end{aligned}$$

of the series (23) converge almost everywhere to $f(x_1, \dots, x_k)$ as n_1, \dots, n_k tend to $+\infty$ independently of one another. The theorem that follows is intermediate between the latter result and the extension of Theorem 1 to the case of k variables.

THEOREM 5. *Let $f(x_1, x_2, \dots, x_k)$ be a function of $k \geq 2$ variables, of period 2π with respect to each x . Suppose that r is an integer satisfying $0 \leq r \leq k-1$, and that the function $|f|(\log^+ |f|)^r$ is integrable. Let $s = k - r$, let $\lambda \geq 1$ be fixed, and let $n_1(t), n_2(t), \dots, n_s(t)$ be non-negative, non-decreasing, integer-valued functions of t tending to $+\infty$ with t . Then at almost every point the means $\sigma_{\mu_1, \dots, \mu_r, \nu_1, \dots, \nu_s}(x_1, x_2, \dots, x_k; f)$ tend to $f(x_1, x_2, \dots, x_k)$, as $\mu_1, \dots, \mu_r, \nu_1, \dots, \nu_s$ tend to infinity, provided that ν_1, \dots, ν_s tend to infinity in such a way that*

$$\lambda^{-1}n_j(t) \leq \nu_j \leq \lambda n_j(t) \quad (j = 1, 2, \dots, s).$$

The corresponding result for Abel means is also true.

Proof. Given an integrable function $g(x)$ of period 2π , let

$$g_*(x) = \sup_{0 < |h| \leq \pi} |1/h \int_0^h |f(x+u)| du|,$$

and let $\sigma_*(x)$ denote the least upper bound of $|\sigma_n(x)|$, σ_n denoting the $(C, 1)$ means of the Fourier series of f . It is well known (Hardy and Littlewood [1]; Zygmund [6], p. 247-248) that $\sigma_*(x) \leq A g_*(x)$. It is also known that if $|g|(\log^+ |g|)^\alpha$ is integrable, so is $g_*(\log^+ g_*)^{\alpha-1}$ and

$$(24) \quad \int_0^{2\pi} g_*(\log^+ g_*)^{\alpha-1} dx \leq A_\alpha \int_0^{2\pi} |g|(\log^+ |g|)^\alpha dx + A_\alpha \quad (\alpha \geq 1)$$

where A_α denotes a constant depending on α only.

Let $\sigma_{\mu_1 \dots \mu_r \nu_1 \dots \nu_s}^*(x_1, x_2, \dots, x_k)$ denote the least upper bound of the numbers $\sigma_{\mu_1 \dots \mu_r \nu_1 \dots \nu_s}(x_1, \dots, x_k)$ under the conditions on $\mu_1, \mu_2, \dots, \mu_r, \nu_1, \dots, \nu_s$ expressed in the statement of Theorem 5. It is enough to prove that for every p , $0 < p < 1$, we have

$$(25) \quad \int_{-\pi}^{\pi} \dots \int_{-\pi}^{\pi} dx_1 \dots dx_r \left\{ \int_{-\pi}^{\pi} \dots \int_{-\pi}^{\pi} \sigma_{\mu_1 \dots \mu_r \nu_1 \dots \nu_s}^* dx_{r+1} \dots dx_k \right\}^{1/p} \\ \leq \frac{A_r \lambda^s}{1-p} \int_{-\pi}^{\pi} \dots \int_{-\pi}^{\pi} |f|(\log^+ |f|)^r dx_1 \dots dx_k + \frac{A_r \lambda^s}{1-p}.$$

That this inequality implies the boundedness almost everywhere of the numbers $\sigma_{\mu_1 \dots \mu_r \nu_1 \dots \nu_s}(x_1, \dots, x_k)$ is clear. In order to prove that it also implies $\sigma_{\mu_1 \dots \mu_r \nu_1 \dots \nu_s}(x_1, \dots, x_k) \rightarrow f(x_1, \dots, x_k)$ almost everywhere, we proceed in a familiar way. First of all, we fix p and apply (25) to the function Mf , where M is a positive constant so large that in the resulting inequality

$$(26) \quad \int_{-\pi}^{\pi} \dots \int_{-\pi}^{\pi} dx_1 \dots dx_r \left\{ \int_{-\pi}^{\pi} \dots \int_{-\pi}^{\pi} \sigma_{\mu_1 \dots \mu_r \nu_1 \dots \nu_s}^* dx_{r+1} \dots dx_k \right\}^{1/p} \\ \leq \frac{A_r \lambda^s}{1-p} \int_{-\pi}^{\pi} \dots \int_{-\pi}^{\pi} |f|(\log^+ |Mf|)^r dx_1 \dots dx_k + \frac{A_r \lambda^s}{M(1-p)}$$

the last term $A_r \lambda^s / M(1-p)$ on the right is $< \epsilon/2$. Then we make a decomposition $f = f' + f''$, where f' is a trigonometric polynomial in x_1, \dots, x_k , and the first term on the right of (25), with f replaced by f'' , is also $< \epsilon$. The final steps may be left to the reader.

Let us revert to (25). Let us fix x_2, \dots, x_k , and let $f_1(x_1, x_2, \dots, x_k)$

be obtained from $f(x_1, x_2, \dots, x_k)$ in the same way as $g_*(x_1)$ is obtained from $g(x_1)$. Similarly we obtain $f_2(x_1, \dots, x_k)$ from $f_1(x_1, \dots, x_k)$, this time fixing x_1, x_3, \dots, x_k , and so on. Denoting for simplicity integrals extended over the k dimensional cell $Q(|x_j| \leq \pi, j=1, 2, \dots, k)$, by

$$\int_{Q_k} \dots d\omega_k, \text{ we get from (24)}$$

$$(27) \quad \int_{Q_k} f_r d\omega_k \leq A \int_{Q_k} f_{r-1} \log^+ f_{r-1} d\omega_k + A \leq \dots$$

$$\leq A_r \int_{Q_k} |f| (\log^+ |f|)^r d\omega_k + A_r,$$

so that f_r is integrable. Let us now observe that

$$|\sigma_{\mu_1 \dots \nu_s, \nu_1 \dots \nu_s}(x_1, \dots, x_r, y_1, \dots, y_s)|$$

$$\leq \pi^{-k} \int_{Q_k} |f(x_1 + u_1, \dots, y_s + v_s)| \prod_{i=1}^r K_{\mu_i}(u_i) \prod_{j=1}^s K_{\nu_j}(v_j) d\omega_k$$

$$\leq A \int_{Q_{k-1}} f_1(x_1, x_2 + u_2, \dots, y_s + v_s) \prod_{i=2}^r K_{\mu_i}(u_i) \prod_{j=1}^s K_{\nu_j}(v_j) d\omega_{k-1}$$

$$\dots$$

$$\leq A_r \int_{Q_s} f_r(x_1, \dots, x_r, y_1 + v_1, \dots, y_s + v_s) \prod_{j=1}^s K_{\nu_j}(v_j) d\omega_s.$$

The last integral, multiplied by π^{-k} , represents the arithmetic mean of the Fourier series of the function $f_r(x_1, \dots, x_r, y_1, \dots, y_s)$ of the variables y_1, \dots, y_s (so that x_1, \dots, x_r are constants). Using the analogue of Theorem 4 (inequality (21)) for the s dimensional case and F' absolutely continuous, integrating the result with respect to x_1, \dots, x_r , and applying (27), we get (25). This completes the proof of Theorem 5.

Remarks. a) Inequality (25) implies that under the assumptions of Theorem 5,

$$\int_{-\pi}^{\pi} \dots \int_{-\pi}^{\pi} dx_1 \dots dx_r \left\{ \int_{-\pi}^{\pi} \dots \int_{-\pi}^{\pi} |\sigma_{\mu_1 \dots \nu_s}(x_1, \dots, x_k) - f(x_1, \dots, x_k)|^p dx_{r+1} \dots dx_k \right\}^{1/p} \rightarrow 0$$

and *a fortiori*,

$$\int_{Q_k} |\sigma_{\mu_1 \dots \nu_s}(x_1, \dots, x_k) - f(x_1, \dots, x_k)|^p d\omega_k \rightarrow 0.$$

b) The part of Theorem 5 pertaining to Abel summability, is imme-

diately extensible to non-tangential paths, as in Theorem B (see Marcinkiewicz and Zygmund [3]) and Theorem 2. Remark a) applies also to that case.

c) The results hold for fractional summability $(C, \alpha_1, \dots, \alpha_k)$ if all the α_j are positive.

d) In Theorem 5 we actually have $r+1$ indices tending to $+\infty$ independently of one another. This is easily seen if one of the functions $n_1(t), \dots, n_s(t)$ is taken as a new independent variable, instead of t .

UNIVERSITY OF CHICAGO.

REFERENCES

- [1] G. H. Hardy and J. E. Littlewood, "A maximal theorem with function-theoretic applications," *Acta Mathematica*, vol. 54 (1930), pp. 81-110.
- [2] B. Jessen, J. Marcinkiewicz and A. Zygmund, "Note on the differentiability of multiple integrals," *Fundamenta Mathematicae*, vol. 25 (1935), pp. 217-234.
- [3] J. Marcinkiewicz and A. Zygmund, "On the summability of double Fourier series," *Fundamenta Mathematicae*, vol. 32 (1939), pp. 122-132.
- [4] C. N. Moore, "On the summability of double Fourier series of discontinuous functions," *Mathematische Annalen*, vol. 74 (1913), pp. 555-578.
- [5] S. Saks, "On the strong derivatives of functions of intervals," *Fundamenta Mathematicae*, vol. 25 (1935), pp. 235-252.
- [6] A. Zygmund, *Trigonometrical series*, Warsaw, 1935.
- [7] A. Zygmund, "On the differentiability of multiple integrals," *Fundamenta Mathematicae*, vol. 23 (1934), pp. 143-149.

ON RIESZ SUMMABILITY AND SUMMABILITY BY DIRICHLET'S SERIES.*

By C. T. RAJAGOPAL.

Addendum and Corrigendum.

Dr. L. S. Bosanquet has pointed out to me that Corollary 2 (This Journal, vol. 59, pp. 374-5) is incorrectly stated. The corollary is a generalization of only the sufficiency part of Schnee's theorem and should run:

Sufficient conditions for the $R(\lambda_n, k)$ -summability ($k \geq 0$) of $\sum_{\nu=1}^{\infty} a_{\nu}$ to s are (i) and (ii'). If $F(t)$ converges for $t > 0$, the conditions $\lim_{t \rightarrow +0} F(t) = s$, $B_k(x) = o(x^{k+1})$ as $x \rightarrow \infty$, are necessary for the $R(\lambda_n, k)$ -summability ($k > 0$) of $\sum_{\nu=1}^{\infty} a_{\nu}$.

Dr. Bosanquet has also kindly suggested the following as a generalization of Schnee's theorem.

COROLLARY 3. *Necessary and sufficient conditions for the $R(\lambda_n, k)$ -summability ($k \geq 0$) of $\sum_{\nu=1}^{\infty} a_{\nu}$ to s are*

(i*) $\frac{t^{k+1}}{\Gamma(k+1)} \int_0^{\infty} A_k(u) e^{-ut} du$ converges (absolutely) for $t > 0$ and tends to s as $t \rightarrow +\infty$.

(ii*) $B_k(x) = o(x^{k+1}), \quad x \rightarrow \infty.$

Further, in the special case in which k is an integer, (ii*) may be replaced by

(ii**) $B_k(x) = o(\lambda_n x^k), \quad \lambda_n \leq x < \lambda_{n+1}.$

Proof. For any $k \geq 0$, the necessity of (i*) is easily proved and that of (ii) follows from the definition of $B_k(x)$ and the fact that both $A_k(x)/x^k$ and $A_{k+1}(x)/x^{k+1}$ tend to s as $x \rightarrow \infty$.

To prove the sufficiency of the conditions (i*) and (ii*), for any $k \geq 0$, we note that (i*) implies the validity of Lemma 4 and that of the proof in 1.2. Consequently (i*) leads to $A_{k+1}(x)/x^{k+1} \rightarrow s$ and thence, in conjunction with (ii*), to $A_k(x)/x^k \rightarrow s$.

The replacement of (ii*) by (ii**), when k is integral, is justified by the fact that the latter condition implies the former whether k is integral or not, while the former implies the latter in the special case of integral k as we can see by putting $\mu = k$, $p = 1$ and taking o instead of O in the following theorem of Dr. L. S. Bosanquet which is to be published in the *Journal of the London Mathematical Society*:

If $B_k(x) = O(x^{k+p})$, where k is a positive integer and $k + p \geq 0$, then, for $\mu = 0, 1, \dots, k$,

$$B_\mu(x) = O\left\{x^\mu \lambda_n^p \left(\frac{\lambda_{n+1}}{\lambda_{n+1} - \lambda_n}\right)^{k-\mu}\right\}, \quad \lambda_n \leq x < \lambda_{n+1}.$$

It may be observed that the special case of Corollary 3 reduces to Schnee's theorem in the usual form when $k = 0$; also that, in any case, we can obtain a variant of the corollary with the integral in (i*) replaced by

$$F_k(t) = t \int_0^\infty \frac{A_k(u)}{u^k} e^{-ut} du.$$

I take this opportunity to draw the attention of the reader to a misprint in condition (ii) of Corollary 1, p. 374. The restriction on $B_k(x)$ should be

$$B_k(x) \geq -K\lambda_n^{k+1}, \text{ not } B_k(x) = -k\lambda_n^{k+1}.$$

MADRAS CHRISTIAN COLLEGE,
TAMBARAM, S. INDIA.

ON MÖBIUS' INVERSION.*

By PHILIP HARTMAN and AUREL WINTNER.

1. Let $x = (x(1), x(2), \dots)$ and $y = (y(1), y(2), \dots)$ be two vectors with an infinity of components. Let E and M denote the infinite matrices of the linear substitutions

$$(1) \quad E: \sum_{m=1}^{\infty} x(nm) = y(n), \quad (n = 1, 2, \dots),$$

and

$$(2) \quad M: \sum_{m=1}^{\infty} \mu(m)y(nm) = x(n), \quad (n = 1, 2, \dots),$$

respectively, where nm or mn denotes the product of m and n , and $\mu(m)$ is the Möbius factor.

Möbius' formal rule for the inversion of the Eratosthenian sieve process states that the linear substitutions (1), (2) are inverses of each other. Correspondingly, if (1) is written as $Ex = y$ and (2) as $My = x$, then $x = My$ will be called the Möbius solution of $Ex = y$, where y is given and x is unknown, and a corresponding manner of speaking will be used if E and M , and x and y , are interchanged.

In a paper appearing in vol. 68 (1946), pp. 321-339, of this Journal, the legitimacy of Möbius' formal inversion was considered. It was shown that if a vector $z = (z(1), z(2), \dots)$ is called *regular* when

$$\sum_{n=1}^{\infty} |z(n)| < \infty,$$

then neither of the Möbius inversions need be legitimate if nothing but the regularity of the respective *solution* vectors, $z = x$ or $z = y$, is assumed; cf. theorems (IV) and (VII), *loc. cit.* On the other hand, various sufficient criteria were developed under which the Möbius inversions are legitimate. The present paper contains further results in the latter direction.

2. The above definition of the "regularity" of a vector is justified by the following fact:

(i) For a given vector y , the system $Ex = y$ cannot have more than one regular solution $x = x_y$.

* Received February 13, 1947.

This is theorem (V), *loc. cit.*, where it is explained that this statement is equivalent to a result of Haar concerning the homogeneous system $Ex = 0$.

It should be emphasized that only the regularity of x is assumed in (i), whereas the given vector y can be arbitrary.

As to the dual of (i), viz., the statement which results if $Ex = y$ is replaced by its formal inverse $My = x$, the situation is as follows: *Loc. cit.*, p. 335, a partial dual of (i) was verified, and the question as to its complete dual was left as a desideratum. It will now be shown that the complete dual of (i) is true:

(ii) *For a given vector x , the system $My = x$ cannot have more than one regular solution $y = y_x$.*

In (ii), no restriction is placed on the given vector x .

A corollary of (ii) is the following dual of Haar's result: *$y = 0$ is the only regular solution of the homogenous Möbius equations $My = 0$.* Needless to say, this corollary of (ii) contains (ii) itself.

The idea of the proof of (ii) is somewhat similar to that of Haar's dual of the last italicized statement (that is, of the fact that $x = 0$ is the only regular solution of the homogeneous Eratosthenian equations $Ex = 0$). The formal details turn out to be more elaborate than in Haar's case.

3. In view of the negative results (IV) and (VII), proved *loc. cit.*, there is a problem concerning the *consistency* of regular solutions and Möbius solutions when both exist. Without any restriction, this question will be left undecided, as it was *loc. cit.* However, some information can be obtained in this regard by the method proving (ii). In fact, a slight modification of the procedure, which would prove (ii) directly, also leads to the following theorem:

(ii*) *For a given regular x , the system $My = x$ cannot have a regular solution y distinct from the Möbius vector Ex , and the latter represents a solution y of $My = x$ whenever the vector Ex is regular.*

In other words, if x is a given regular vector, then, according as the vector Ex is or is not regular, the system $My = x$ has the unique regular solution $y = Ex$ or no regular solution at all. The point in this alternative is the fact, mentioned before, that the concepts of Möbius solutions and regular solutions are, in general, distinct. The italicized statement following (ii), which is equivalent to (ii), makes it particularly clear that (ii*) is a refinement of (ii); so that only (ii*) will have to be proved.

There arises the question whether or not (i) can be refined to a theorem,

say (i*), in the same way as (ii) is refined to (ii*). The answer turns out to be affirmative:

(i*) For a given regular y , the system $Ex = y$ cannot have a regular solution x distinct from the Möbius vector My , and the latter represents a solution x of $Ex = y$ whenever the vector My is regular.

The proof of (i*) will depend on an adaptation of Haar's proof for the uniqueness of the regular solution of the homogeneous system $Ex = 0$.

It is worth mentioning that (i*) and (ii*), respectively, make clear the methodical rôle of the steps by means of which (VIII) and (IX) have been verified *loc. cit.*

4. In order to prove (ii*), suppose that x is a given regular vector with reference to which the system $My = x$ has at least one regular solution, y . It will be shown that this y must then be the vector Ex . This will imply the first assertion of (ii*).

It will be sufficient to prove that $y(1)$, the first component of the vector y , is the first component of the vector Ex ; i. e., that (1) must then hold for $n = 1$. In fact, suppose that this has been deduced from the system (2). Then it can be applied to the system which results from the full system (2) when n , in (2), is restricted to multiples of a fixed positive integer, say of i ; that is, to the system

$$\sum_{m=1}^{\infty} \mu(m)y(mni) = x(ni), \quad (n = 1, 2, \dots).$$

Since the system which corresponds to the latter system in the same way as (1) corresponds to (2) is

$$(3) \quad \sum_{m=1}^{\infty} x(mni) = y(ni), \quad (n = 1, 2, \dots),$$

and since the assertion is supposed to be true for $n = 1$, it follows that

$$\sum_{m=1}^{\infty} x(mi) = y(i).$$

Since this means that (1) itself is true for $n = i$ and, since i is arbitrary, it is seen that it is sufficient to prove (1) for $n = 1$.

Let j and k be two positive integers and let the equations (2) be summed over those values of the index n which are composed of the first j primes $p_1 = 2, p_2 = 3, \dots$ with multiplicities not exceeding k (the value $n = 1$ is included). This gives

$$(4) \quad \sum_{h_1=0}^k \cdots \sum_{h_j=0}^k x(p_1^{h_1} \cdots p_j^{h_j}) = \sum_{h_1=0}^k \cdots \sum_{h_j=0}^k \sum_{m=1}^{\infty} \mu(m) y(m p_1^{h_1} \cdots p_j^{h_j}).$$

Since the vector $y = (y(1), y(2), \cdots)$ is supposed to be regular, and since $\mu(m)$ is a bounded function of m , the expression on the right of (4) is absolutely convergent. It can be rearranged into

$$(5) \quad \sum_{m=1}^{\infty} \sum_{d|m} \mu(m/d) y(m),$$

if the index of the interior summation is restricted to those divisors d of m which satisfy the following restrictions:

$$(6) \quad d = p_1^{h_1} \cdots p_j^{h_j} \text{ and } 0 \leq h_i \leq k, \quad (i = 1, 2, \cdots, j).$$

If d runs through all divisors of m , then, by the definition of the Möbius function, the sum

$$(7) \quad \sum_{d|m} \mu(m/d)$$

is 1 or 0 according as $m = 1$ or $m > 1$. On the other hand, $\mu(n)$ is $(-1)^g$ or 0 according as n is the product of g distinct primes or is not square-free. Hence, it is easily verified that, if the summation index is restricted to those divisors d of m which are enumerated under (6), then the corresponding sum (7) is

$$(8) \quad \mu(m), \quad (-1)^j \mu(m/P^{k+1}) \text{ or } 0$$

according as m is relatively prime to the product $P = P_j = p_1 \cdots p_j$, the quotient m/P^{k+1} is an integer relatively prime to P or m is in neither of these cases; that is, according as

$$(9) \quad (m, P) = 1, \quad m = lP^{k+1} \text{ and } (l, P) = 1 \text{ or } *$$

where the asterisk denotes the negation of the first two cases of (9) and the three cases of (9) correspond to the respective cases of (8).

It follows that, if d is restricted as in (6), then the double sum (5) is identical with

$$(10) \quad \sum_{(m, P)=1} \mu(m) y(m) + \sum_{(l, P)=1} \mu(l) y(lP^{k+1}).$$

Since the expression on the left of (4) is identical with the sum (5) restricted by (6), which is the expression (10), and since $|\mu(l)| \leq 1$, it follows that the absolute value of the difference

$$(11) \quad \sum_{h_1=0}^k \cdots \sum_{h_j=0}^k x(p_1^{h_1} \cdots p_j^{h_j}) - \sum_{(m, P)=1} \mu(m) y(m)$$

cannot exceed

$$(12) \quad \sum_{(l, P)=1} |y(lP^{k+1})|.$$

Hence, (11) is majorized by

$$(13) \quad \sum_{l=p^{k+1}}^{\infty} |y(l)|,$$

(13) being a majorant of (12).

The indices k and j , hence $P = p_1 \cdots p_j$, had fixed values thus far. Now let $k \rightarrow \infty$, while j is fixed. Then (13), hence (11), tends to 0. In view of the regularity of the vector x , this result can be written in the form

$$\sum_{h_1=0}^{\infty} \cdots \sum_{h_j=0}^{\infty} x(p_1^{h_1} \cdots p_j^{h_j}) = \sum_{(m,P)=1}^{\infty} \mu(m)y(m).$$

Since $\mu(1) = 1$ and $|\mu(m)| \leq 1$, this implies that

$$(14) \quad \left| \sum_{h_1=0}^{\infty} \cdots \sum_{h_j=0}^{\infty} x(p_1^{h_1} \cdots p_j^{h_j}) - y(1) \right| \leq \sum_{m=p_j+1}^{\infty} |y(m)|.$$

Finally, let $j \rightarrow \infty$. Then, since the vector y is supposed to be regular, the expression on the right of (14) tends to 0. On the other hand, since the vector x is supposed to be regular, it is clear that the difference, the absolute value of which occurs on the left of (14), tends to

$$\sum_{m=1}^{\infty} x(m) - y(1),$$

as $j \rightarrow \infty$. Hence, the latter expression is 0. This proves that (1) is true for $n = 1$.

5. In order to complete the proof of (ii*), the truth of the following assertion remains to be verified:

(ii bis) *If x and Ex are regular vectors, then $y = Ex$ is a solution y of the system $My = x$.*

In view of the definitions, (1) and (2), of E and M , the assertion (ii bis) can be formulated as follows: If

$$(15) \quad \sum_{n=1}^{\infty} |x_n| < \infty \quad \text{and} \quad \sum_{n=1}^{\infty} \left| \sum_{m=1}^{\infty} x_{mn} \right| < \infty,$$

then

$$(16) \quad \sum_{n=1}^{\infty} \mu(n) \sum_{m=1}^{\infty} x(mni) = x(i)$$

holds for $i = 1, 2, \dots$. Corresponding to the above reduction of the case on arbitrary n in (1) to the case $n = 1$ of (1), it is sufficient to show that (15) implies

$$(17) \quad \sum_{n=1}^{\infty} \mu(n) \sum_{m=1}^{\infty} x(mn) = x(1),$$

which is the case $i = 1$ of (16); in other words, that (17) holds when both x and Ex are regular.

Instead of considering, as above, two independent indices, k and j , consider only one of them, j , and put again $P = p_1 \cdots p_j$. Then, in view of the first of the assumptions (15),

$$(18) \quad \sum_{h_1=0}^1 \cdots \sum_{h_j=0}^1 \mu(p_1^{h_1} \cdots p_j^{h_j}) \sum_{m=1}^{\infty} x(mp_1^{h_1} \cdots p_j^{h_j}) = \sum_{(m,P)=1} x(m)$$

is a "logical identity" (Sylvester), implied by the definition of the Möbius factor. If $j \rightarrow \infty$, then the convergence of the first of the series (15) implies that the sum on the right of (18) tends to $x(1)$. Hence, in order to prove (17), it is sufficient to show that the expression on the left of (18) tends to the expression on the left of (17), as $j \rightarrow \infty$. But this becomes clear from the second of the assumptions in (15) if it is observed that, on the one hand, $\mu(n)$ is a bounded function of n and, on the other hand, $\mu(n)$ is 0 when n is not square-free.

This completes the proof of both (ii bis) and (ii*).

6. In order to prove (i*), suppose that y is a given regular vector with reference to which the system $Ex = y$ has at least one regular solution, x . It will be shown that this x must then be the vector My . This will imply the first assertion of (i*).

The claim is that (2) holds for every n . Corresponding to the above reductions of an arbitrary n to the case of $n = 1$, it is sufficient to show that (2) holds for $n = 1$.

To this end, choose an arbitrary integer j , put $P = P_j = p_1 \cdots p_j$, multiply the n -th of the equations (1) by the factor $\mu(n)$ and sum the result over those values of n which are divisors of P . This gives

$$\sum_{h_1=0}^1 \cdots \sum_{h_j=0}^1 \mu(p_1^{h_1} \cdots p_j^{h_j}) y(p_1^{h_1} \cdots p_j^{h_j}) = \sum_{(m,P)=1} x(m).$$

The case $n = 1$ of (2) now follows in the same way as (ii bis) did, the last formula line playing the part of (18).

In order to complete the proof of (i*), the truth of the following dual of (ii bis) remains to be verified:

(i bis) *If y and My are regular vectors, then $x = My$ is a solution x of the system $Ex = y$.*

The proof of this fact can be omitted, since it differs from the proof of the first part of (ii*) only in obvious details.

DESARGUES' AND PAPPUS' GRAPHS AND THEIR GROUPS.*

By I. N. KAGNO.

1. Introduction. The graph Δ consisting of the vertices $A, B, C, D, E, F, G, H, I, J$, and the arcs $AC, AD, AG, AH, AI, AJ, BD, BE, BF, BH, BI, BJ, CE, CF, CG, CI, CJ, DF, DG, DH, DJ, EF, EG, EH, EI, FG, FJ, GH, HI, IJ$, is called Desargues' Graph. The graph Π consisting of the vertices $A, B, C, D, E, F, G, H, I$, and the arcs $AD, AE, AF, AG, AH, AI, BD, BE, BF, BG, BH, BI, CD, CE, CF, CG, CH, CI, DG, DH, DI, EG, EH, EI, FG, FH, FI$, is called Pappus' graph. These graphs are defined by Sister Van Straten in an abstract of a paper to be published,¹ and she states the Theorems, (1) Desargues' Graph is an irreducible non-toroidal graph; (2) Pappus' Graph can be imbedded in a torus. The purpose of this note is to present a few additional properties of these graphs, and to derive their groups.

In a previous paper² we have defined the group \mathcal{G} of a graph G as follows; Corresponding to any one-one continuous map of G into itself there is a substitution τ on its vertices a_1, \dots, a_n . Corresponding to the group of all possible maps of G into itself there is a substitution group \mathcal{G} on the letters a_1, \dots, a_n . \mathcal{G} is called the group of G , and we say that G has the group \mathcal{G} . In this paper we also defined the adjacency number I_b^a of a pair of vertices a, b of G as follows: If G contains the arc ab , then $I_b^a = I_a^b = 1$. If G does not contain ab , then $I_b^a = I_a^b = 0$.

2. Desargues' graph.

THEOREM 1. *Desargues' Graph Δ cannot be imbedded in a projective plane.*

* Received November 12, 1946.

¹ Sister Petronia Van Straten, "Toroidal and non-toroidal graphs," *Bulletin of the American Mathematical Society*, vol. 52 (1946), p. 831, Abstract No. 345. She uses a numeral notation to designate the vertices instead of the notation we find it convenient to use here. The full paper is to appear in *Reports of a Mathematical Colloquium* (Notre Dame), No. 8.

² "Linear graphs of degree ≤ 6 and their groups," *American Journal of Mathematics*, vol. 68 (1946), pp. 505-520. We shall refer to this paper as [L]. In this paper the author gave an example of a graph of Degree 7, with no vertex of degree < 3 , having no non-identical substitution group. It is of interest to note that if the restriction on the degree of the vertices is removed it is possible to find a graph of fewer than seven vertices which has no non-identical substitution group. Namely, the graph consisting of the vertices a, b, c, d, e, f , and the arcs $ab, bc, bd, ce, de, cf, ef$, has this property. (We omit the proof, which can readily be supplied by the reader of [L].)

Proof. Δ contains as subgraph the graph Δ^* consisting of the vertices A, \dots, J , and the arcs $AC, AD, AI, AJ, CF, DF, CI, CJ, DH, HI, BD, BJ, IJ, BE, EH, EF$. Δ^* is homeomorphic with the irreducible non-projective-planar graph G_{4v-a-2} discussed by the author in a previous paper.³ Hence Δ is non-projective-planar.

Definition. Let G be a connected graph of n vertices, having no simple loops. Let N be the complete n -point formed by joining each pair of non-adjacent vertices of G by an arc, and let $G' = N -$ the arcs of G . G' will be called the *complement* of G .

THEOREM 2. *The complement of Desargues' Graph is Petersen's Graph.*

Proof. The complement of Δ is the graph Δ' consisting of the vertices of Δ and the arcs $AB, AE, AF, BC, BG, CD, CH, DE, DI, EJ, FH, FI, GI, GJ, HJ$. Let us replace the letters A, \dots, J by unordered number couples as follows; $A = (1, 2)$, $B = (4, 5)$, $C = (2, 3)$, $D = (1, 5)$, $E = (3, 4)$, $F = (3, 5)$, $G = (1, 3)$, $H = (1, 4)$, $I = (2, 4)$, $J = (2, 5)$. Then it will be seen that Δ' is precisely Petersen's Graph.⁴

LEMMA. *If G' is the complement of G , then G and G' have the same group.*

Proof. Let p, q be any pair of vertices. If $I_q^p = 1$, ($= 0$), in G , then in G' , $I_q^p = 0$, ($= 1$). Now suppose τ is any map of G into itself. Since τ is arc preserving, $I_q^p = I_{\tau(q)}^{\tau(p)}$. Now since in G' , I_q^p and $I_{\tau(q)}^{\tau(p)}$ each have the opposite value they have in G , $I_q^p = I_{\tau(q)}^{\tau(p)}$ in G' also. Consequently τ maps G' into itself.⁵ Conversely any map of G' into itself also maps G into itself. Hence G and G' have the same groups.

THEOREM 3. *Desargues' Graph has a group of degree ten and order 120, which is simply isomorphic with the symmetric group \mathfrak{S}_5 on five letters.*

Proof. By the lemma, Δ has the same group as its complement, Petersen's Graph. But Petersen's Graph has the group given in the Theorem.⁶

³ "The mapping of graphs on surfaces," *Journal of Mathematics and Physics*, vol. 16 (1937), pp. 46-75; page 66 and plate I on page 62. Note that on page 66 the symbol for G_{4v-a-2} is misprinted. It should read $G_{4v-a-2} = (B/0/p/pp_{1d}, pp_{d0}, pp_{d0})$.

⁴ R. Frucht, "Die Gruppe des Petersenschen Graphen . . ." *Commentarii Mathematici Helvetici*, vol. 9 (1937), pp. 217-223.

⁵ [L], Theorem 1.

⁶ Frucht, *loc. cit.* The group of Δ can be obtained from \mathfrak{S}_5 as follows; Replace the letters A, \dots, J by the number couples as was done in the proof of Theorem 2. Every substitution of \mathfrak{S}_5 on the numbers 1, . . . , 5 will determine a substitution on the distinct number couples, that is, a substitution on the letters A, \dots, J . \mathfrak{S}_5 can be gene-

3. Pappus' graph.

THEOREM 4. *Pappus' Graph Π cannot be imbedded in a projective plane.*

Proof. Π contains as subgraph the graph Π^* consisting of the vertices of Π and the arcs $AD, AE, AF, BD, BE, BF, CD, CE, CF, GD, GE, GF, HD, HE, HF$. Π^* is homeomorphic with the irreducible non-projective-planar graph G_{111-0} .⁷ Hence Π is non-projective-planar.

THEOREM 5. *Pappus' Graph has the group*

$\mathfrak{G} = \mathfrak{G}_1 \times \mathfrak{G}_2 \times \mathfrak{G}_3 \times \mathfrak{G}_4 = (ABC)all(DEF)all(GHI)all(ADG \cdot BEH \cdot CFI)$
of order 1296.⁸

Proof. We note the following adjacency numbers for Π

$$I_B^A = I_C^B = I_C^A = 0, \quad I_B^D = I_F^E = I_F^D = 0, \quad I_H^G = I_I^H = I_I^G = 0,$$

and $I_v^v = 1$ for any other pair of vertices of Π .

A set of generators for \mathfrak{G} is $[(ABC), (AB), (DEF), (DE), (GHI), (GH), (ADG)(BEH)(CFI), (AD)(BE)(CF)]$. It can readily be seen that each of these substitutions maps Π into itself. Hence Π is mapped into itself by every substitution of \mathfrak{G} .

Suppose Π is mapped into itself by a substitution

$$\tau = \begin{pmatrix} A & B & C & D & E & F & G & H & I \\ a & b & c & d & e & f & g & h & i \end{pmatrix} \neq 1. \quad \text{We shall show that } \tau \in \mathfrak{G}.$$

Case 1. $\tau(A)$ is A, B , or C . Since $I_A^B = 0, I_A^C = 0, I_a^p = 1$, (where $a = A, B$, or C ; $p \neq A, B$, or C), τ cannot carry B or C into D, E, F, G, H , or I . That is, a, b, c is a permutation of A, B, C . The subgroup \mathfrak{G}_1 of \mathfrak{G} contains a substitution σ such that $\sigma^{-1} = \begin{pmatrix} a & b & c \\ A & B & C \end{pmatrix}$. Let

$$\alpha = \sigma^{-1}\tau = \begin{pmatrix} A & B & C & D & E & F & G & H & I \\ A & B & C & d & e & f & g & h & i \end{pmatrix}.$$

(a) If $d \neq G, H$, or I , since $I_D^m = 0, I_d^q = 1$, (where $m = E$ or F ; $q = G, H$, or I ; $d = D, E$, or F), α cannot carry E or F into G, H , or I . That is, d, e, f is a permutation of D, E, F . The subgroup \mathfrak{G}_2 of \mathfrak{G} contains a substitution ρ such that $\rho^{-1} = \begin{pmatrix} d & e & f \\ D & E & F \end{pmatrix}$, and the subgroup \mathfrak{G}_3 contains a

rated by the substitutions (1234) and (15). These correspond to the substitutions $(AHEC)(BFJD)(GI)$ and $(AJ)(BH)(FG)$ respectively, which generate the group of Δ .

⁷ "The mapping of graphs . . ." *loc. cit.*, page 65 and plate I on page 62.

⁸ Where $\mathfrak{G}_4 = \{1, (ADG)(BEH)(CFI), (AGD)(BHE)(CIF), (AD)(BF)(CF), (AG)(BH)(CI), (DG)(EH)(FI)\}$.

substitution μ such that $\mu^{-1} = \begin{pmatrix} g & h & i \\ G & H & I \end{pmatrix}$. Hence $\mu^{-1}\rho^{-1}\sigma^{-1}\tau = 1$ and $\tau = \sigma\rho\mu \in \mathfrak{G}$.

(b) If $d = G, H$, or I , since $I_D^n = 0$, $I_d^n = 1$, (where $n = E$ or F ; $\tau = D, E$, or F ; $d = G, H$, or I), α cannot carry E or F into D, E , or F . That is, d, e, f is a permutation of G, H, I , and consequently g, h, i is a permutation of D, E, F . The subgroup \mathfrak{G}_3 of \mathfrak{G} contains a substitution λ_1 such that $\lambda_1^{-1} = \begin{pmatrix} d & e & f \\ G & H & I \end{pmatrix}$, and the subgroup \mathfrak{G}_2 contains a substitution λ_2 such that $\lambda_2^{-1} = \begin{pmatrix} g & h & i \\ D & E & F \end{pmatrix}$. Hence $\lambda_1^{-1}\lambda_2^{-1}\alpha = \begin{pmatrix} A & B & C & D & E & F & G & H & I \\ A & B & C & G & H & I & D & E & F \end{pmatrix}$. The subgroup \mathfrak{G}_4 contains the substitution $\lambda_3 = (DG)(EH)(FI)$. Hence $\lambda_3\lambda_1^{-1}\lambda_2^{-1}\alpha = \lambda_3\lambda_1^{-1}\lambda_2^{-1}\sigma^{-1}\tau = 1$, and $\tau = \sigma\lambda_2\lambda_1\lambda_3 \in \mathfrak{G}$.

Case 2. $\tau(A)$ is D, E , or F . Since $I_A^k = 0$, $I_a^k = 1$, (where $k = B$ or C ; $a = D, E$ or F ; $s = A, B, C, G, H$, or I), τ cannot carry B or C into A, B, C, G, H or I . That is, a, b, c is a permutation of D, E, F . The subgroup \mathfrak{G}_2 of \mathfrak{G} contains a substitution ν such that $\nu^{-1} = \begin{pmatrix} a & b & c \\ D & E & F \end{pmatrix}$. Let $\beta = \nu^{-1}\tau = \begin{pmatrix} A & B & C & D & E & F & G & H & I \\ D & E & F & d & e & f & g & h & i \end{pmatrix}$.

(a) If $d = A, B$, or C , since $I_D^l = 0$, $I_d^l = 1$, (where $l = E$ or F ; $t = D, E, F, G, H$, or I ; $d = A, B$, or C), β cannot carry E or F into D, E, F, G, H , or I . That is, d, e, f is a permutation of A, B, C . Now \mathfrak{G}_1 contains a substitution ϕ_1 such that $\phi_1^{-1} = \begin{pmatrix} d & e & f \\ A & B & C \end{pmatrix}$, and \mathfrak{G}_3 contains a substitution ϕ_2 such that $\phi_2^{-1} = \begin{pmatrix} g & h & i \\ G & H & I \end{pmatrix}$, and hence $\phi_2^{-1}\phi_1^{-1}\beta = \begin{pmatrix} A & B & C & D & E & F \\ D & E & F & A & B & C \end{pmatrix}$. \mathfrak{G}_4 contains a substitution $\phi_3 = (AD)(BE)(CF)$. Hence $\phi_3\phi_2^{-1}\phi_1^{-1}\beta = \phi_3\phi_2^{-1}\phi_1^{-1}\nu^{-1}\tau = 1$, and $\tau = \nu\phi_1\phi_2\phi_3 \in \mathfrak{G}$.

(b) If $d = G, H$, or I , since $I_D^v = 0$, $I_d^v = 1$ (where $v = E$ or F ; $u = A, B, C, D, E$, or F ; $d = G, H$, or I), β cannot carry E or F into A, B, C, D, E , or F . That is, d, e, f is a permutation of G, H, I , and consequently g, h, i is a permutation of A, B, C . Now \mathfrak{G}_3 contains a substitution θ_1 such that $\theta_1^{-1} = \begin{pmatrix} d & e & f \\ G & H & I \end{pmatrix}$, and \mathfrak{G}_1 contains a substitution θ_2 such that $\theta_2^{-1} = \begin{pmatrix} g & h & i \\ A & B & C \end{pmatrix}$ and hence $\theta_2^{-1}\theta_1^{-1}\beta = \begin{pmatrix} A & B & C & D & E & F & G & H & I \\ D & E & F & G & H & I & A & B & C \end{pmatrix}$. \mathfrak{G}_4 contains a substitution θ_3 such that $\theta_3^{-1} = (ADG)(BEH)(CFI)$. Hence $\theta_3^{-1}\theta_2^{-1}\theta_1^{-1}\beta = \theta_3^{-1}\theta_2^{-1}\theta_1^{-1}\nu^{-1}\tau = 1$ and $\tau = \nu\theta_1\theta_2\theta_3 \in \mathfrak{G}$.

Case 3. $\tau(A)$ is G, H , or I . The proof is similar to that of Case 2.

NEW YORK CITY.

THE ISOPERIMETRIC PROBLEM IN THE MINKOWSKI PLANE.*

By HERBERT BUSEMANN.

The slow progress in the theory of Finsler spaces as compared to Riemann spaces is partly due to lack of information regarding the corresponding local, that is the Minkowskian, geometry. Those Minkowskian features will contribute most to an understanding of Finsler spaces which are not merely *verbal* generalizations of *known* euclidean statements.¹ The purpose of the present note was originally only to show that the isoperimetric problem (for any dimension) in Minkowski spaces leads to such a feature.

It turned out, however, that the plane problem can be solved in a general form—no longer significant for Finsler spaces—and then exhibits a phenomenon which is of interest for the theory of isoperimetric problems in the calculus of variations. The result seems to indicate that the standard methods may have followed too closely the pattern of the fixed endpoint problem. For that reason the plane case is here presented separately.

The following are the results: let $F(x, y)$ be continuous, positive for $x, y \neq 0$, and positive homogeneous of order 1. The problem, to find among all simple closed curves $x(t), y(t)$ with a given orientation and a given Minkowski length $L = \int F(\dot{x}, \dot{y}) dt$ one which bounds the greatest (euclidean) area, *has a unique solution* (up to translations) *no matter whether the indicatrix $C : \mathfrak{L}(x, y) = 1$ is convex or not.* For non-convex C the solution is the same as for the boundary \bar{C} of the convex closure of C as indicatrix and is *homothetic to the polar reciprocal* (figuratrix) *of \bar{C} with respect to the unit circle rotated through $\pm \pi/2$.*

For Finsler spaces only the case is of interest where C is convex and has the origin as center. Since Finsler or Minkowski area differs from the euclidean area by a constant factor, *the solution of the Minkowskian isoperimetric problem is the same as for the above problem.* In intrinsic Minkowskian terms it may be described as the curve of length L for which as new

* Received January 22, 1947.

¹ Since the euclidean geometry is a special Minkowskian geometry, every theorem on general Minkowskian spaces is a generalization of some euclidean fact.—It has been stated that the so-called relative differential geometry is the differential geometry of Minkowski space. This is however not so. Relative length or area are not Minkowski length or area.

unit circle (indicatrix) perpendicularity (transversality) is the reverse of perpendicularity with respect to C . It is in general not a Minkowski circle.²

The result for non-convex C suggests the question: what is the relation between the lengths defined by C and \bar{C} as indicatrices respectively. An answer is given in the final section for general n -dimensional Finsler spaces to the effect that *the corresponding Lebesgue lengths are equal*.

1. The form and the uniqueness of the solution for a *convex indicatrix* C can be derived in a few lines from the *Brunn-Minkowski Theory*. If $x = r \cos \theta$, $y = r \sin \theta$ and $F(\cos \theta, \sin \theta) = \rho(\theta)$, then $r = \rho^{-1}(\theta)$ is the equation of C in polar coordinates.

Let D be an analytic closed convex curve which contains the origin in its interior. If $h(x, y)$ is the supporting function of D then $h(\theta) = h(\cos \theta, \sin \theta)$ is the distance of the origin from the tangent of D at the point q where the exterior normal of D has direction θ . The radius of curvature of D at q is $h(\theta) + h''(\theta)$ (see [1, p. 65]) so that the euclidean line element of D at q equals $(h(\theta) + h''(\theta))d\theta$. Therefore the Minkowski length $L(D)$ of D is

$$(1) \quad L(D) = \int_0^{2\pi} (h(\theta) + h''(\theta)) \rho(\theta + \delta\pi/2) d\theta$$

where $\delta = 1$ ($\delta = -1$) if the orientation of D is positive (negative). From now on we fix the orientation of D and give δ the corresponding value.

If C is convex then $F(x, y)$ is convex, hence $H(x, y) = r\rho(\theta + \delta\pi/2)$ is convex and therefore the supporting function of a certain convex curve K . Then $L(D)$ is by (1) twice the mixed area $A(D, K)$ of D and K (see [4, p. 245]),

$$(2) \quad L(D) = 2A(D, K).$$

Both $L(D)$ and $A(D, K)$ do not change when D or K undergo translations ([1, p. 40]) and depend continuously on D . Since every convex curve can be approximated by analytic convex curves ([1, section 27]), the relation (2) holds for any convex curve D . The Theorem of Brunn-Minkowski yields ([1, p. 97]) that

$$(3) \quad A^2(D, K) \geq A(D)A(K),$$

² This fact was first noticed by S. Ulam who found, and communicated to the author, that the solution for $F(x, y) = \max(|x|, |y|)$ is $|x| + |y| = \text{const.}$ In this connection it is of interest that *the Minkowski circle is the solution of the problem to find among all curves with a given Minkowski diameter one which bounds the greatest area*, compare [2, section 2].

where $A(D)$ and $A(K)$ are the areas bounded by D and K respectively, and the equality sign holds only when D and K are homothetic. The relations (2) and (3) yield the isoperimetric inequality

$$(4) \quad L^2(D)/A(D) \geq 4A(K).$$

If D is a general simple closed curve with the given orientation, then $L(D)$ is to be defined by the Weierstrass sum. If D is not rectifiable in the euclidean sense, then $L(D) = \infty$. If D is rectifiable and suitably parametrized (for instance by the arc length) then the Weierstrass sum coincides with the Lebesgue integral $\int F(\dot{x}, \dot{y}) dt$ along D (see [5, p. 51]). If C is convex the euclidean straight lines are shortest connections. Hence, if D is not convex the boundary of the convex closure of D has by the usual arguments at most the length of D and bounds a greater area, so that (4) holds for any simple closed curve with the given orientation, and the equality still holds only for D homothetic to K .

It is well known that the curve \bar{K} with the supporting function $F(x, y)$ (the so called figuratrix of F) originates from C by a polar reciprocity with respect to the euclidean unit circle (see [4, pp. 146, 147]). Then K is obtained from \bar{K} by a rotation through $-\delta\pi/2$ about O .

2. This way of generating K leads to the property of K which permits the discussion of *non-convex indicatrices*. Call tangent of K a supporting line of K which is at the same time a right hand or left hand tangent of K at one of its common points with K . Then (if C is convex)

- (5) No tangent of K is parallel to a radius of C from O to an interior point of a (euclidean) segment ab which lies on C .

For the polars of the different points of ab with respect to $x^2 + y^2 = 1$ pass through one point p and are supporting lines of \bar{K} at p . Hence a supporting line \bar{S}_x of \bar{K} which corresponds to an interior point x of ab cannot be a tangent of \bar{K} . After rotation of \bar{K} through $-\delta\pi/2$ this fact becomes (5).

Let now C be not convex and call $\bar{F}(x, y)$ the positive, positive homogeneous function of degree 1 for which $\bar{F}(x, y) = 1$ is the boundary \bar{C} of the convex closure of C . Then $\bar{F}(x, y) \leq F(x, y)$ and if $\bar{L}(D)$ denotes the length of D measured by \bar{F} ,

$$(6) \quad \bar{L}(D) \leq L(D).$$

If K denotes the curve originating from \bar{C} by a polar reciprocity in $x^2 + y^2 = 1$ and rotation through $-\delta\pi/2$, then by (4) and (6)

$$L^2(D)/A(D) \geq \bar{L}^2(D)/A(D) \geq 4A(K)$$

and the equality signs can hold only if D is homothetic to K and $L(D) = \bar{L}(D)$. The last relation holds always for D homothetic to K .

For if $u(t), v(t)$ represents K with the euclidean arc length t as parameter, then u and v exist except for an at most countable number of t 's and

$$L(D) = \int F(\dot{u}, \dot{v}) dt, \quad \bar{L}(D) = \int \bar{F}(\dot{u}, \dot{v}) dt.$$

If $F(x, y) > \bar{F}(x, y)$, then the ray from O to (x, y) intersects C in a point where C has no supporting line, and therefore \bar{C} in an interior point of a segment ab which is part of \bar{C} . The statement (5) shows now that $F(u, v) = \bar{F}(\dot{u}, \dot{v})$ for any t for which \dot{u} and \dot{v} exist, which proves $L(D) = \bar{L}(D)$. The following has been proved

THEOREM 1. *Let $F(x, y)$ be continuous, positive for $x, y \neq 0$ and positive homogeneous of degree 1. If $r = \bar{\rho}^{-1}(\theta)$ is the polar equation of the boundary \bar{C} of the convex closure of $F(x, y) = 1$, then for any simple closed curve D*

$$(7) \quad L^2(D)/A(D) \geq 2 \int_0^{2\pi} (\bar{\rho}^2(\theta) - \bar{\rho}'^2(\theta)) d\theta$$

where $L(D)$ is length in terms of F and $A(D)$ is the euclidean area.

The equality sign holds for positively (negatively) traversed D only if D is homothetic to the polar reciprocal K of C with respect to $x^2 + y^2 = 1$ rotated about O through $-\pi/2$ ($\pi/2$).

The right side of (7) equals $4A(K)$ by a known integral representation of area in terms of the supporting function, see [1, p. 66]. It is true that this reference states the formula only for analytic curves, but its validity for general convex curves follows from the fact that if the analytic convex curves $r = \rho_v^{-1}(\theta)$ tend to $r = \rho(\theta)$, then their derivatives ρ'_v tend automatically to ρ' , see [1, p. 35].

3. Area is defined only for symmetric Finsler spaces [see 2, section 2]. Therefore, as far as Minkowski geometry is concerned, Theorem 1 is interesting only when C is convex and has O as center ($F(x, y) = F(-x, -y)$). In that case the (non-oriented euclidean straight) line g is said to be perpendicular to the line h (or h transversal to g) for the Minkowski geometry with unit circle C , in a formula: $g \perp_C h$, if the parallel to h through O intersects C in those points where C has supporting lines parallel to h . This terminology is due to the fact that the intersection of g and h minimizes the Minkowski

distance of a given point on g from a variable point on h . If K is defined as before, then the statements

$$(8) \quad g \perp_O h \text{ and } h \perp_K g \text{ are equivalent}$$

This may be seen as follows: If p is a point of C and h is a supporting line of C at p let the euclidean normal to h from the origin O intersect h at f and \bar{K} at \bar{p} . By the definition of \bar{K} the line Op must intersect a supporting line \bar{h} of \bar{K} at \bar{p} at a right angle. Denote the intersection by \bar{f} . If \bar{p} , \bar{f} , \bar{h} , are transformed into p_1 , f_1 , h_1 after rotation through $-\pi/2$ about O , then Op is parallel to h_1 and Op_1 is parallel to h . This means: if $Op = g$ is an arbitrary radius of C and h is a supporting line of C at p then the supporting line h_1 of K parallel to g contains a point p_1 for which $g_1 = Op_1$ is parallel to h , which is equivalent to (8).

Minkowski area (measure) is the euclidean area (Lebesgue measure) multiplied by π and divided by the euclidean area $A(C)$. Hence, if $M(D)$ denotes the Minkowski area bounded by D , then

$$M(D) = 2\pi A(D) / \int_0^{2\pi} \rho^{-2}(\theta) d\theta,$$

where $r = \rho^{-1}(\theta)$ is the polar equation of C . The relation (7) yields then the *Minkowskian Isoperimetric Inequality*

$$(9) \quad L^2(D)/M(D) \geq \pi^{-1} \int_0^{2\pi} (\rho^2 - \rho'^2) d\theta \int_0^{2\pi} \rho^{-2} d\theta,$$

and the equality sign holds only for the curves K which satisfy (8).

The right side of (9) is a Minkowski invariant, hence it does not change if the x, y undergo a non-degenerate affine transformation.

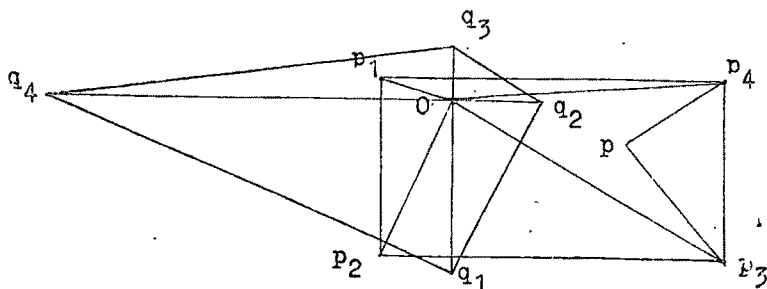
For the theory of Finsler spaces it is interesting to find out in which cases the Minkowski circles solve the isoperimetric problem. By (8) the symmetry of the relation $g \perp_O h$ is necessary and sufficient. The analytic equivalent to this statement is that the supporting function $h(\theta)$ of C must be proportional to $\rho(\theta + \pi/2)$. The factor of proportionality is easily evaluated by expressing $A(C)$ first in terms of the radius vector and then in terms of the supporting function of C . It is then found that

$$(10) \quad h(\theta) = \left[\int_0^{2\pi} \rho^{-2}(\theta) d\theta \right]^{1/2} \cdot \left[\int_0^{2\pi} (\rho^2(\theta) - \rho'^2(\theta)) d\theta \right]^{-1/2} \rho(\theta + \pi/2).$$

All curves with a continuous non-vanishing curvature (the Legendre condition of the calculus of variations) which satisfy (10) have been

determined by Radon in [6]. But there are other curves with this property, for instance the regular n -gons with $n \equiv 2 \pmod{4}$.³

To have examples for the preceding theory it is useful to construct K for any convex polygon C , not necessarily with O as center. Since the convex closure of any polygon is again bounded by a polygon, this construction will yield K for arbitrary polygonal C . To obtain the counterclockwise solution corresponding to C let p_1, p_2, \dots, p_n be the vertices in counterclockwise order (see figure). Orient the radii



of C from p_i to O . Then K is the polygon q_1, \dots, q_n such that the oriented side $q_{i-1}q_i$ is parallel to p_iO and Oq_i is parallel to p_ip_{i+1} (subscripts are to be reduced mod n if necessary). In the figure the polygon K is also the solution for the polygon $C' = (p_1p_2p_3pp_4)$ which has C as boundary of its convex closure. If ab denotes also the euclidean distance of the points a, b and β_i the angle $p_{i-1}Op_i$, the following formula determines the lengths of the sides of K :

$$(11) \quad \frac{L(K)}{2A(K)} = \frac{1}{b^{1-1}b} \frac{Op_{i-1}^{-1} \sin \beta_{i+1} - Op_i^{-1} \sin (\beta_i + \beta_{i+1}) + Op_{i+1}^{-1} \sin \beta_i}{\sin \beta_i \sin \beta_{i+1}}.$$

This relation can be obtained through solving the isoperimetric problem for the polygon p_1, \dots, p_n directly by means of elementary calculus and trigonometry. (11) contains the inequality (3), including the discussion of the equality sign, and yields thus a new elementary proof for the Brunn-Minkowski inequality for polygons. By a limit process the inequality for general convex curves is obtained, but the "only if" part in the condition for the equality sign is lost. For a regular n -gon (11) yields

³ While this paper was in print a geometric construction for all curves C with symmetric $\perp c$ was given by M. M. Day in: "Some characterizations of inner-product spaces," *Transactions of the American Mathematical Society*, vol. 62 (1947), pp. 320-337. The author seems to be unaware of the earlier work of Radon, Helly, and Blaschke on symmetric transversality.

$$L^2(D)/M(D) \geq 4\pi^{-1}n^2 \sin^2(\pi/n)$$

which becomes for $n \rightarrow \infty$ the ordinary isoperimetric inequality.

4. The discussion in Section 2 leads to the question: what is for arbitrary D the relation between $L(D)$ and $\bar{L}(D)$. This question will be answered completely in the present section for general n -dimensional Finsler spaces because it throws some light on the implications of the Lebesgue type of definition of length and area.

Let $F(x_1, \dots, x_n; \alpha^1, \dots, \alpha^n) = F(x, \alpha)$ be defined and continuous for all contravariant vectors (x, α) of an n -dimensional manifold of class 1, positive for $\alpha \neq 0$, and positive homogeneous of order 1 in the α^i . The Lebesgue F -length of a continuous curve $x(t)$ is defined as

$$\lambda_{\bar{F}}^L(x(t)) = \inf [\liminf \int F(x_\nu(t), \dot{x}_\nu(t)) dt]$$

for all sequences $\{x_\nu(t)\}$ of curves of class D' which tend to $x(t)$ in the sense of Fréchet.

With $F(x, \alpha)$ we associate a quasiregular integrand $\bar{F}(x, \alpha)$ as follows.⁴ For fixed x (in a definite coordinate neighborhood) let \bar{C}_x be the boundary of the convex closure of the indicatrix $F(x, \alpha) = 1$ in α -space. Then $\bar{F}(x, \alpha)$ is the integrand whose indicatrix (in the same x -system) is \bar{C}_x . The relation between $L(D)$ and $\bar{L}(D)$ is then determined by the following general theorem.

THEOREM 2. *For any continuous curve $x(t)$*

$$\lambda_{\bar{F}}^L(x(t)) = \lambda_{\bar{F}}^L(x(t)).$$

Since \bar{F} is quasiregular the ordinary \bar{F} -length $\lambda_{\bar{F}}^L(x(t))$ of a curve⁵ coincides with the Lebesgue length $\lambda_{\bar{F}}^L(x(t))$, so that Theorem 2 may also be formulated as

THEOREM 2'. *For any continuous curve $x(t)$*

$$\lambda_{\bar{F}}^L(x(t)) = \lambda_{\bar{F}}(x(t)).$$

⁴ For further details see [3]. The following considerations follow closely Section 3 of that paper. The present Theorem 2 is a generalization of Theorem 1, [3, p. 184].

⁵ This means again the limit of the Weierstrass sum $\sum \bar{F}(x(t_i), x(t_{i+1}) - x(t_i))$ which for absolutely continuous $x(t)$ coincides with the Lebesgue integral $\int \bar{F}(x, \dot{x}) dt$. The F -length λ_F is defined correspondingly. Regarding these questions compare [5, p. 51].

Proof. Since $\bar{F}(x, \alpha) \leq F(x, \alpha)$ and $\lambda_{\bar{F}}$ is lower semicontinuous it follows that for a suitable sequence of curves $x_\nu(t)$ of class D' which tend to x

$$\lambda_{\bar{F}}^L(x) = \lim \lambda_{\bar{F}}^L(x_\nu) \geq \liminf \lambda_{\bar{F}}^L(x_\nu) \geq \lambda_{\bar{F}}^L(x).$$

Since $\lambda_{\bar{F}}^L$ is lower semicontinuous it suffices to prove $\lambda_{\bar{F}}^L(x) \geq \lambda_F^L(x)$ for x which are of class D' . For then with a suitable sequence of curves of class D' which tend to x

$$\lambda_{\bar{F}}^L(x) = \lim \lambda_{\bar{F}}^L(x_\nu) \geq \liminf \lambda_F^L(x_\nu) \geq \lambda_F^L(x).$$

Since both lengths are additive it is also sufficient to consider a curve $x(t)$, $a \leq t \leq b$, of class C' which lies entirely in one coordinate neighborhood. If the norm of the partition $[t] = [t_0 = a < t_1 < t_2 < \dots < t_{N+1} = b]$ tends to 0

$$\Sigma \bar{F}(x(t_i), x(t_{i+1}) - x(t_i)) \rightarrow \lambda_{\bar{F}}(x).$$

The direction of α is called quasiregular for $\bar{F}(x, \alpha)$ if $F(x, \alpha)$ possesses at its point of the form $\delta\alpha$, $\delta > 0$, a supporting plane. The relation $F(x, \alpha) = \bar{F}(x, \alpha)$ holds if and only if α is quasiregular. If α is not quasiregular choose $\delta > 0$ such that $\delta\alpha$ lies on $\bar{F}(x, \alpha) = 1$. Then $m, 2 \leq m \leq n$, regular directions β_1, \dots, β_m exist such that

$$(12) \quad \delta\alpha = \Sigma \delta_j \beta_j, \quad \delta_j > 0, \quad \Sigma \delta_j = 1$$

For $F(x, \alpha) = 1$ is connected, hence n points β_1, \dots, β_n on $F(x, \alpha) = 1$ exist such that the $(n-1)$ -simplex spanned by β_1, \dots, β_n contains $\delta\alpha$ (see [1, p. 9]). If $\delta\alpha$ is not in the interior of the simplex choose the notation so that the $(m-1)$ -simplex spanned by β_1, \dots, β_m contains $\delta\alpha$ in its interior. There is a hyperplane through $\delta\alpha$ which is simultaneously a supporting plane P of $F(x, \alpha) = 1$ and $\bar{F}(x, \alpha) = 1$ (see [1, p. 6]). Then P cannot separate any two of the points β_1, \dots, β_n and must therefore contain all the points β_1, \dots, β_m , so that all these points are regular. This proves (12).

Applying this result to $\alpha_i = x(t_{i+1}) - x(t_i)$ yields relations of the form

$$\alpha_i = \sum_j \delta_{ij} \beta_j, \quad \delta_{ij} > 0.$$

Because F is homogeneous and the simplex spanned by $\beta_{i1}, \dots, \beta_{im_i}$ lies by construction on $\bar{F}(x, \alpha) = 1$,

$$\bar{F}(x(t_i), x_i) = \sum_j \bar{F}(x(t_i), \delta_{ij}, \beta_{ij}) = \sum_k \bar{F}(x(t_i), y_{ik+1} - y_{ik})$$

where y_{ik} is the point $x(t_i) + \sum_{j=1}^k \delta_{ij} \beta_{ij}$. Since all the β_{ik} are quasiregular

$$F(x(t_i), y_{ik+1} - y_{ik}) = \bar{F}(x(t_i), y_{ik+1} - y_{ik}).$$

if $p(\cdot)$ is the euclidean polygon with vertices

$$y_0 = x(a), y_{01}, \dots, y_{0m_0} = x(t_1), y_1, \dots, y_{1m_1} = x(t_2), \dots, y_{nm_N} = x(b)$$

then $\int F(p, \dot{p}) dt$ will, for sufficiently small norm of $[t]$ differ arbitrarily little from

$$\sum_i \sum_k \bar{F}(x(t_i), y_{ik+1} - y_{ik}) = \sum_i \bar{F}(x(t_i), x(t_{i+1}) - x(t_i)).$$

Hence, as the norm of $[t]$ tends to 0, the curve $p(t)$ tends to $x(t)$ in the sense of Fréchet and

$$\int F(p, \dot{p}) dt \rightarrow \int \bar{F}(x, \dot{x}) dt$$

which proves the theorem, because $\lambda_{\bar{F}}^L(x) \leq \lim \int F(p, \dot{p}) dt$.

A consequence of Theorem 2, which is contained in the results of [3] and which can be generalized to planes as minimal surfaces in Minkowski spaces is this:

For any integrand $F(\alpha)$ in a Minkowski space the straight lines are shortest connections of their endpoints for Lebesgue F -length.

This is true since they are shortest connections for \bar{F} -length.

UNIVERSITY OF SOUTHERN CALIFORNIA.

REFERENCES

- [1] T. Bonnesen and W. Fenchel, "Theorie der konvexen Körper," *Ergebnisse der Mathematik*, vol. 3, part 1, Berlin, 1934.
- [2] H. Busemann, "Intrinsic area," *Annals of Mathematics*, vol. 48 (1947), pp. 234-267.
- [3] H. Busemann and W. Mayer, "On the foundations of calculus of variations," *Transactions of the American Mathematical Society*, vol. 49 (1941), pp. 173-198.
- [4] H. Minkowski, *Gesammelte Abhandlungen*, vol. II, Leipzig, 1911.
- [5] C. Ponce, "La méthode métrique en calcul des variations," *Actualités Scientifiques et Industrielles*, No. 885, Paris, 1941.
- [6] F. Radon, "Ueber eine besondere Art ebener konvexer Kurven," *Berichte über die Verhandlungen der Königlich Sächsischen Gesellschaft der Wissenschaften zu Leipzig*, vol. 68 (1916), pp. 123-168.

CORRECTIONS.

By I. N. KAGNO.

The author wishes to make the following corrections to his paper "Linear Graphs of Degree ≤ 6 and their Groups," this Journal, vol. LXVIII, pp. 505-520.

Page 507, line 28 should read

"Suppose that $I_{a_1, a_2}^{a_1} = 1, \dots$ ".

Page 514, line 14 should read

$H_5 \equiv [ab, ad, ae, af, bc, be, bf, cd, ce, cf, de, df]$.

Page 514, Theorem 3.4 should read

H_5 has the group $\mathfrak{S}_{32} \equiv (abcdef)_{48}$.

Page 520, Theorem D.12 should read

$\mathfrak{S}_{32} \equiv$ has the graph H_5 (Theorem 3.4).

Page 520, delete lines 20 and 21.

Page 520, line 10 should read

The group \mathfrak{S}_{16} has the graph H_4 . (Theorem 3.3).
